DRAFT FOR COMMENTS

WILLINGNESS TO PAY FOR NATIONAL FRESHWATER QUALITY IMPROVEMENTS

by

Robert Cameron Mitchell

and

Richard T. Carson

Quality of the Environment Division

Resources for the Future
1755 Massachusetts Ave. N.W.
Washington, D.C.

USEPA GRANT # CR810224-01-4

Dr. Alan Carlin
Project Officer
Office of Policy Analysis
U.S. EnvironmentalProtection Agency
Washington, D.C. 20460

October 1984

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Table of Contents

EXECUTIVE SUMMARY

This report presents the findings of a study of the national benefits of freshwater quality Improvements. The objectives of the study were to conduct exploratory research to determin the applicability of the contingent valuation method to measuring national freshwater benefits. The research on which the report Is baaed occurred in two stages. During the first stage we engaged In extensive instrument development and undertook a large pilot study. The field work for the pilot study occurred In 1980 and the findings were reported in 1981 (Mitchell and Carson, 1981). Our work during the second stage, Involved a series of pretests In which the Instrument was further refined, and a field study in which a national probability sample of 814 people were interviewed In person by professional Interviewers. The data reported here come from l heae interviews which were conducted by the Opinion Research Corporation of Princeton, New Jersey under our direction. Chapter 4 describes the development of the research instrument and the text of the instrument is presented in appendix A.

'We value people's willingness to pay for three levels of national freshwater quality which we defined as : Boatablee-- where virtually all the nation's freshwater lakes, rivers and stream are boatable and many have higher water quality; Fishable -- where virtually all are fishable and some have higher water quality; and Swimmable -- where virtually all are swimmable. These levels, which our pretests showed were meaningful to the respondents and which correspond with the national water quality goals, were described in words and depicted on a water quality ladder which we developed for this study. The consumres WTP amounts include both recreational and intrinsic benefits.

Respondents expressed their willingness to pay by means of a payment vehicle of annual taxes and higher prices. They were reminded that they are already paying for this program In their current taxes and prices, although they were not informed of the amount until later in the interview after they gave their flrat WTP amounts. An anchored payment card elicitation technique was used In this study in lieu of the bidding game. This procedure was developed for this study. Chapter S contains a justification for the method and presents the results of experiments we conducted to determine the degree to which it Is vulnerable to bias. We conclude that the anchored payment card represents a significant Improvement over the widely used bidding game method and, for this study at least, constitutes a viable and useful elicitation procedure.

Four different willingness to pay (WTP) amounts were measured during the course of the Interview. After respondents gave an Initial amount, they were offered the chance to revise it. The revised amount (WTP_R) Is the basis for our lower bound estimate of water quality benefits. The third amount was obtained after informing respondents the amount households of their Income category are already paying for national water quality improvemnts. In order to test for bias Induced by respondents giving us amounts for environmental Improvements more generally (policy part-whole bias, a subtype of amenity misspecification), half the sample was also told what they are paying for air pollution control. The resulting informmed (WTP_I) amount Includes any revisions the respondents wished to make after receiving this information. The final amount #a obtained after respondents were asked if they would **Increase** their UT? amounts If they were not enough to reach any of the three goals. This constitutes an upper bound on our estimates.

The instrument also obtained information about whether the respondents

would still be willing to pay their revised amount for swimmable quality water "if the beat we could do #a to raise the minimum only halfway from fishable to swimmable." This question was asked of half the sample while the other half was aaked the 95 percent question, which asked them if they would still be willing to pay the fishable amount "if five percent of the nation's water bodfea remain at the boatable level..." Respondents also divided their WTP_R amounts between their states and the rest of the nation. Additional Information was obtainned on their recreational use of water, their attitudes towards environmental issues, and a wide range of background variables.

Of the original 813 interviews, 564 or 70 percent yielded usable WTP amounts. In order to minimize item nonresponse and sample selection bias, we imputed WTP values for the thirty percent with missing WTP values using CART, a tree structured classification program. We then uaed household weights supplied by the Opinion Research Corporation to weight the observations to make the sample representative of the Census population.

Chapter 2, describes the findings In detail. We obtain adjusted annual household values of $99 for boatable quality water, $70 for fishable and $78 for swimmable for a total willingness to pay for national water quality benefits of $242 with a 95 percent confidence Interval of $205-279. These estimates are consistent with those obtained In our 1980 pilot study, thua showing stability. Evidence for their reliability and (construct) validity is suggested by our l atlmatIon of the log-log form which gives an adjusted R^2 of .36. Chapter 3 discusses a number of other factors relevant to the estimates reliability and validity and concludes that they constitute defensible measures of national freshwater benefits.

Among the other findings are the following: (1) Our test for Policy-package part-whole bias Is negative, it doesnot appear to be present ;(2)

Respondents allocate approximately one third of their benefits for out-of-state water; (3) Many respondents are indifferent to whether a full improvement (e.g. to swimmable) or only a partial Improvement occurs, being willing to pay the same amount for each; (4) The distributional benefits of water quality are mildly progressive as measured by the percent of their income respondents am willing to pay for this prpose.

If we take our adjusted WTP_R total value to be an estimate Of the lower bound for household willingness to pay to achieve a water quality goal of 99 percent swimmable water, an aggregate annual national benefit of $20.3 billion is indicated for possible benefits with a 95 percent confidence interval of $17.0 - $23.5 billion. A possible upper bound is the adjusted WTP amounts given after the respondents were informed what they are currently paying for water quality improvements. This yields an aggregate value of $24.0 billion.

Preface

The particular methodological approach we adopt in this study, a national contingent valuation survey, emerged as we studied the problem of how beat to measure national freshwater quality benefits. It builds on a tradition of innovative research using the contingent valuation methodology which extends back to the 1960s and which has flourished during the 1970s as economists have grappled with the challenging task of measuring benefits.

Several years ago, Robert Havemann, commenting on a paper which analyzed 60 benefit studies, declared: "To me, the situation is...extremely discouraging, because, in my view,what has passsed for benefit estimates In these studies forms a catalog of what not to do in cost-benefit analysis" (Haveman, 1975). In our endeavor to avoid joining this infamous roll of abortive or misguided benefit studies we have attempted to address the issues of reliability and validity in as much detail as possible. To help the reader evaluate the extent to which we have succeeded in this task, we provide as much information as possible in this report about how we developed our Instrument, the decisions we made at various stages in our data analysis, and the patterns of responses. Since a contingent valuation is only as good as its questionnaire and sample, the text of the questionnaire and detailed Information about the sample are presented in two of the appendices.

The report/s structure consists of two parts. In the first we describe the study, present our findings,and discuss their reliability and validity. The second part is devoted to methodological considerations. One chapter describes the evolution of the instrument, a lengthy process the account of which may be useful to others whocontemplate conducting such a study in the future. The

other chapter in this part discusses the anchored payment card elicitation method which we developed for this study.

With the necessary disclaimer that we alone are responsible for the work reported here, we wish to gratefully acknowledge the assistance we have received over the years from our colleagues here at RFF and clamhere. We benefitted from early discussions with Ralph d'Arge and David Brookshire of the University of Wyoming, George Tolley of the University of Chicago, Alan Randall of the University of Kentucky, Richard Bishop and Thomas Heberlein of Wisconsin University, Kerry Smith of Vanderbilt University, William Deavouages and Kirk Pate of the Research Triangle Institute, W. Michael Hanemann of the University of California, Berkeley and Alan Carlin, our project monitor at EPA. At RFF, Raymond Kopp and Michael Hazilla offered us much useful counsel on statistical and econometric problems as has William J. Vaughan. Vaughan also prepared the index for our water quality ladder and helped us refine our theoretical and conceptual ideas. Clifford S. Russell was extremely generous with his time.

Robert Cameron Mitchell

Richard T. Carson

October, 1984

Part I. THE BENEFITS OF CLEAN WATER

1. THE DESIGN OF THE STUDY

Introduction

The question of whether or not the benefits of water quality improvement programs as a whole are greater than or equal to their mats has long been of interest to economists and policy makers (Peskin and Seskin, 1975; Council on Environmental Quality, 1979; Feenberg and Hills, 1980; Tolley, Yaron and Blomquist ,1983). Although numerous water benefit studies have been conducted in the past thirty years, they have been of limited use in estimating national water quality benefits (Tihansky 1975; Freeman 1982).[1] For example, site values derived from travel ooat studies generally do not control for water quality (Dwyer, Kelly and Bowes,1977) nor can they measure nonuae values. Of the handful of studies which directly measure nonuae benefits using contingent valuation (Gramlich, 1977; Oster 1977; Greenley, Walsh and Young, 1981, 1982; Desvousges, Smith and McGivney, 1982; Blomquist, 1983a), all value the benefits of one particular site. or area,making extrapolation to the entire United States problematic.

---

1. Freeman (1982) notes a general tendency for studies which use the appropriate economic theory and suitable estimation teohniquea, such as Feenberg and Mills (1980), to have an inadequate data base for making national estimates.

In this report,we present new estimates of the national benefits of attaining the national freshwater quality goals of boatable, fishable, and swimmable water. In addition to the national benefit estimates, we also assess the benefits of intermediate improvements and the distribution of the benefits by income group. These estimates avoid moat of the difficulties usually associated with extrapolating from local or regional studies to the national level or with aggregating benefi ta across households, spatial areas, or types of benefits because they are baaed on a contingent valuation survey of a national sample of the United States population. This methodology, of course, carries with It its own set of problems which w will address both in this chapter and, especially, in the next which evaluates the reliability and validity of our findings.

The Contingent Valuation Method

Contingent valuation (CV) uses survey research techniques to elicit people's preferences in the form of ulllingneaa-to-pay (WTP) monetary amounts. In fta standard form, the CV survey describes a detailed hypothetical market in which a specified public good may be purchased and asks a respondent how much of their household's current income in dollars they would be willing to give up in exchange for a specified increase in level of the public good. Usually the valuation queatfona are repeated several times for different levels of the good so that a Hicksian compensated demand curve is traced out. Since its initial applications in the 1960s (Davis 1963; Knetsch and Davis 1966; Hammack and Brown,1974) considerable effort has been devoted to establishing its theoretical basis,[2] developing the methodology,[3] and, where possible, comparing

2. For arguments that CV data are generated in forms consistent with the theory of welfare change measurement see Randall, Ives and Eastman (1974), Brookshire, Randall, and Stoll (1980), Just, Hueth and Schmitz (1982) and Deavouagea, Smith and McGivney (1983).

its estimates with those wing market demand-based measures.'

There now appears to be widespread agreement that the correct measure of benefits la willingness-to-pay and that, provided it can be administered without bias, the contingent valuation method can be used to estimate the Hicksian consumer surplus measures (Freeman, 1979; Just, Hueth, and Schmitz, 1982).[5] If one accepts the compensating surplus form of WTP as the appropriate welfare measure for a apeclfled improvement in the water quality enjoyed by an individual household, and takes the current distribution of income as given; then a point on the Samuelson-Bradford bid (or benefit) curve (Bradford, 1970; Raadall et al.,,1974) la given by summing all households' WTP amounts for the new level of water quality. Optimal provision of water quality is the point at which the aggregate marginal coat and benefit curves cross (see Figure 1).

The key problems la measuring WTP by the contingent valuation method now appear to be empirical rather than theoretical. One set of problems involves basic conceptual features of the scenario or description of the hypothetical market. In order to conduot this study, we had to answer questions such as the following: What la the current level of national water quality? This question has never really been answered,but which la one must be addressed since that level la our reference or baseline level of water quality. What property right do individuals have for the current level? The answer to this question bears

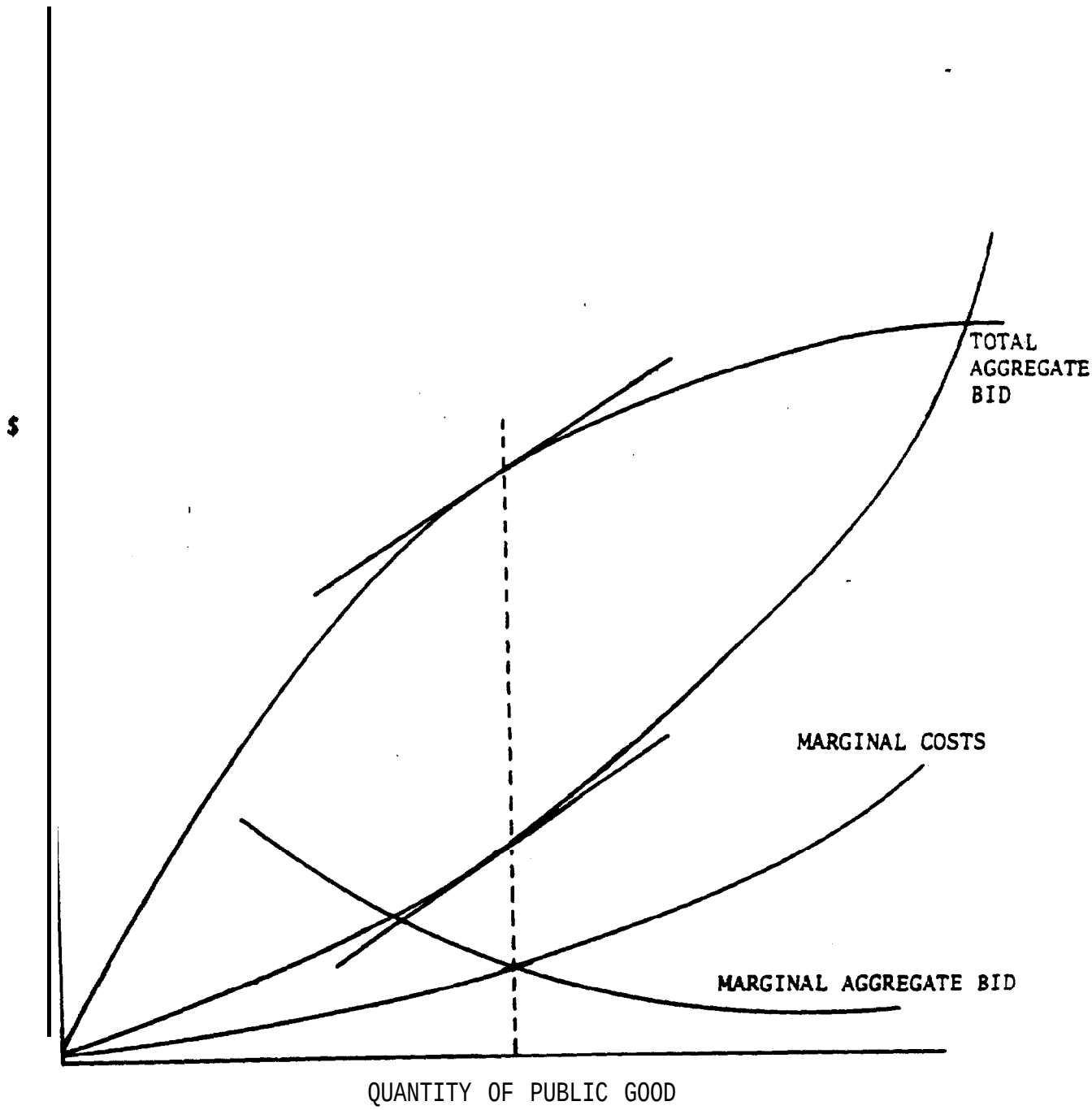
---

3. Cummings et al. (1984) comprises a major review of this work. See also Brookahin and Crocker (1981), Schulze, d'Arge and Brookshire (1981), Randall, Hoehn and Brookshire (1983) and Mitchell and Carson (1984a).

4. Studies making travel cost compulaona are Knetshch and Davis (1966), Bishop and Heberlein (1979), Desvousges, Smith and McGivney (1983) and Seller, Stoll and Chavas (1984). Those making hedonic price comparisons are Brookshire, Thayer, Schulze, and d'Arge (1982), Cummings, Schulze, Brookshire, and Gerking (1983), and Blomquist (1983b).

5. The relationship between welfue economics and contingent valuation is examined in detail in Mitchell and Carson (1984a).

Figure 1 Optimal level of the provision of a public good as a function of marginal cost and marginal aggregate benefits.



From Randall, Ives and Eastman (1974).

importantly on the wording of the elicitation questiona. What levels of water quality are both sufficiently understandable by respondents and meaningful to policy makers to be usefully valued? Exactly which benefits arc measured by the contingent valuation exercise?

Closely related to these are a second set of problems which concern the need to make the survey sufficienty meaningful to respondents so that they understand the qwstions and are motivated to give truthful and considered answers. The act of determining dollar values for a nonmarketed public good such as national water quality is unfamiliar to our respondents and the concept of a minimum national water quality level is difficult to convey to those with lower levels of education. We describe below a number of features of our survey instrument which address the Issues of reliability and validity.

A final set of problems involves the procedures used to analyze the data and generalize from the usable survey responses to the national population. Without a prohibitively expensive number of repeat visits to obtain interviews at the selected households, even the best designed national sample will be unrepresentative of the population in noticeable ways and will require weighting in order to prevent biased national WTP estimates. Compared with standard survey questions, a higher level of item nonresponses to the willingness-to-pay questions is common and to be expected in CV surveys.[6] In addition to weighting the sample to compensate for underrepresented demographic groups, we inpute the missing responses for the WTP question[8] wing CART, a recently developed tree structured classification procedure (Breiman et al.

6. In national surveys, five to fifteen percent nonresponse is normal for questions involving income related items. Since an unusually high level of respondent effort is required to give considered answers to the willingness to pay questions in CV surveys, it is preferable to accept a somewhat higher than normal level of item nonresponse than to have marginal respondents give thoughtless answers to these questions.

1984).

Conceptual Issues

Levels of Water Quality

The Clean Water Act of 1972 and its amendments suggest three levels of minimum national water quality which should be valued: boatable, fishable, and swimmable. In the survey instrument used in this study, we we a **water quality** ladder developed in our 1981 study, and shown in figure 2, as a visual aid. The three levels of water quality were located on the ladder whose top and bottom were defined as the best and worst possible water quality. Matching these levels of water quality with physical' water quality criteria is no easy task nor is there complete agreement on how to do this. Their placement was determined by an index developed by W.J. Vaughan of Resources for the Future. Appendix C discusses the basis on which the ladder was constructed in some detail. This ladder has subsequently been adopted for use in other CV studies (e.g., Desvousges, Smith and McGivney, 1983).

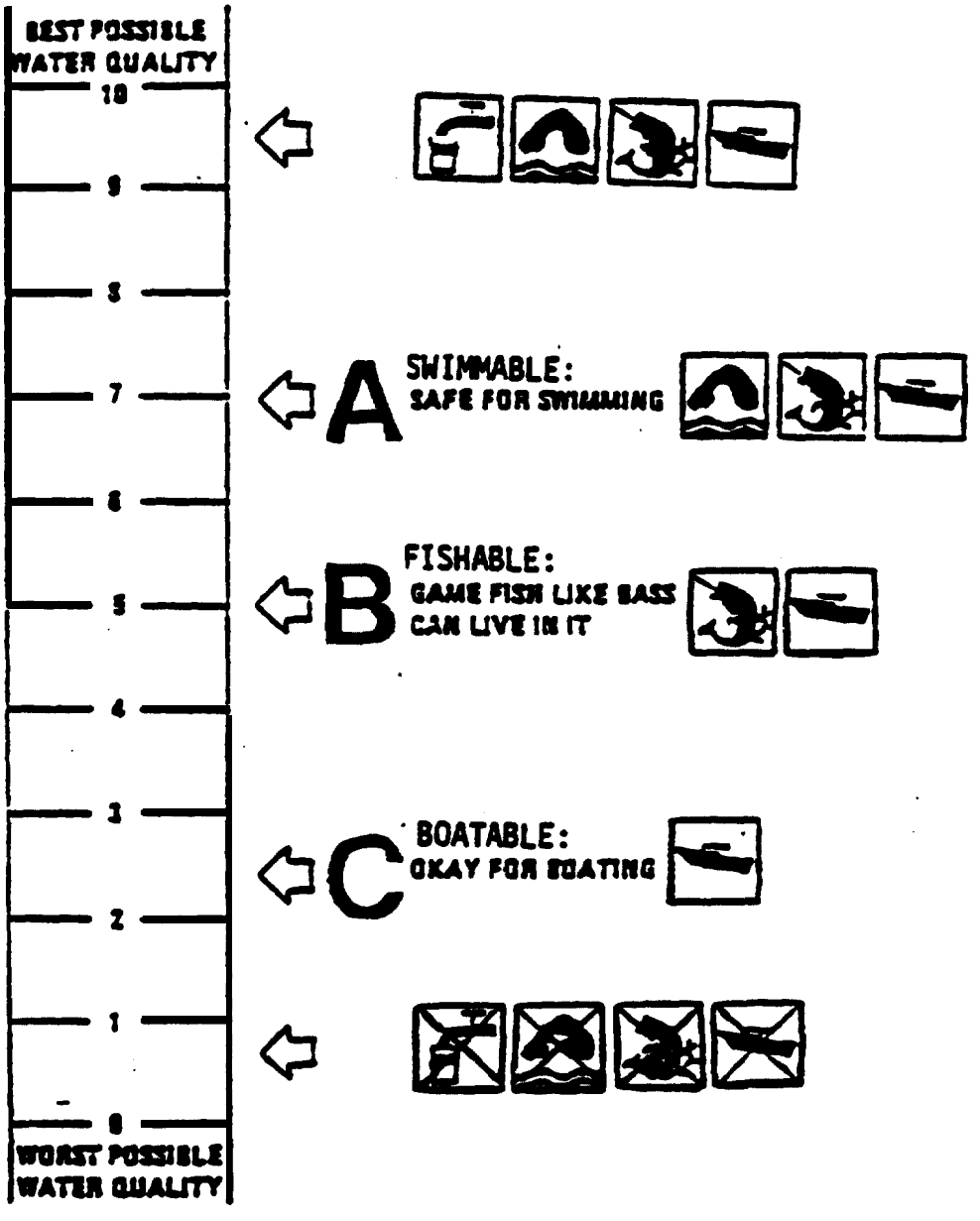
Property Right Structure

Determining the appropriate reference or, since we have adopted a compensating surplus view, the status quo level of water quality, is somewhat more problematic. Depending on how the reference level is defined, the **correct** form of the elicitation question is either: Row much are you willing to pay (WTP) to keep this level or how much are you willing to accept **(WTA)** in compensation for the loss of a given level of quality. The prevailing practice has been to substitute the WTP format for the WTA format wherever the latter has been indicated because the WTA form does not give valid data in practice. Too many respondents react to the notion of selling their right to the amenity by refusing to answer or by giving protest answers such as demanding infinite compensation. Recognition of the measurement problems associated with the

Figure 2

RESOURCES FOR THE FUTURE

WATER QUALITY LADDER



willingness to accept (WTA) measures led CV researchers to propose that only WTP measures should be used (Cummings, et al., 1984). The argument has been made by most CV researchers (e.g., Brookshire, Randall and Stoll, 1980; Mitchell and Carson, 1981; Desvousges, Smith and McGivney, 1983) that where a WTA measure is indicated by theory, it can be replaced by the sore tractable WTP version without bias owing to the Willig (1976) bounds. Hanemann's recent theoretical work suggests that this rationale for substituting a WTP measure for a WTA measure is no longer valid. In this section, we review this development and propose a new approach to conceptualizing the correct Hicksian contingent surplus measures for CV studies which we apply in this study.

According to the Willig bounds, under most conditions the difference between a WTA and a WTP measure of the same good should be inconsequential.[7] CV theorists (Freeman, 1979; Schulze et al., 1981; Brookshire et al., 1982) extrapolated this finding to CV studies and concluded that the substitution of WTP for WTA had little empirical effect. This was a convenient finding, given the methodological problems with WTA. However, Hanneman (1983, 1984a, 1984b), has recently shown that the Willig bounds between compensating variation, ordinary consumer surplus and equivalence variation do not necessarily hold when consumers are only offered discrete choices. His findings are disturbing to those who would substitute WTP formats for WTA format where the latter is indicated, be&se they suggest that there can be very large divergences when

---

7. 4s an illustration of the differences between the consumer surplus measures, considered a person with an Income (Y) of $18,000, a Willingness-to-pay for an increase in the level of provision of the public good in question of $250 and a prioe flexibility of income for the good of 7%. Using equations (11) M - WTP/M M/2Y and (13) WTA - WTP M2 from Randall and Stoll (1980) and the information above, we can solve for l PYA and M Marshallian). The WTP measure is the copensating surplus, WTA is the equivalence surplus, so we have CS ($250) < M ($251.22) < ES ($252.45). Thus the difference between the smallest and largest consumer surplus is approximately one percent or $2.50.

the slopes of the the 'relevant Hlckslan and Marshellian demand curves are not similar. Furthermore, the general tendency of these differences is that WTA > WTP. Since CV studies all involve discrete choices, this means that WTP measures cannot be substituted for an indicated WTA measure without the possibility of seriously biased results. The largest difference between WTA and WTP occurs when the Hicksian equivalence demand curve (WTP) is a straight line shile the Hicksian compensating surplus demand curve is kinked, forming a right angle where the respondent is asked for his or her willingness to accept a reduction in level of the good in question. The more frequently the good is traded the less likely this sltwtlon exists.

To date, the discussion of the correct Hlckslan measure has not differentiated between the property rights implied by private and public goods. A rethinking of the nature of the property right implied in public goods offers a new perspective on the WTA measurement dilemma. Table 1 shows the appropriate Hlckslan measures for private and public goods for different levels of access and ownership. The Hicksian measures for private goods are shown on the left hand side, where the relevant dimensions are use and private ownership. For public goods, shown on the right, the relevant dimensions are availability and individual vs. collective possession.

At the present time the nation's lakes, rivers and streams meet a minimum standard of boatable quality. Among the currently unavailable minimum levels are those where every freshwater body would be at least at the fishable or swimmable levels. The potential for availability is included in our discussion of public goods because they involve various kinds of intrinsic values in addition to we values. For example, a portion of consumers' consumer surplus for water quality may come from the option value consumers place on the knowledge that a given quality level is available for use even though these

Table 1. COMPENSATING SURPLUS MEASURES FOR PRIVATE AND PUBLIC GOODS

	Private			Publics	
	Own	Not Gwn		Individually Held	Collectively Held
Use	%A	ES_{WTP}	Level Available or Potentlally Available	CS_{WTA}	CS_{WTP}
Do Not Use	ES_{WTP}	CS_{WTP}	Level Not Available	CS_{WTP}	CS_{WTP}

CS = Compensating Surplus
WS = Equivalent Surplus

'For public goods which require annual payments (or their equivalents) to maintain a given level of the good.

consumers do not themselves presently use or plan to we it. For private goods where individuals have legally defined exclusive property rights to particular goods, use, not access, is the relevant dimension.

Turning to the tables other axis, the primary distinction of importance for the property rights to private goods is whether the good is owned by a consumer or not. The appropriate CS and ES measures of consumer surplw follow from this determination. In the case of public goods, which are collectively owned, the important determinant of property rights is whether the good is collectively or individually hold.

The first type of property right for public goods we call "individually held. " In this case, individuals are granted exclusive rights to use some public good by the relevant governing body because the granting of such rights is deemed to serve the public interest in some manner. Typically the goods so affected are excludable snd subject to congestion. Various allocation rules

are wed, ranging from auctions to free grants based on principles such as competency and first-come-first-served. Occasionally, as with mining claims on public land and broadcasting frequencies, the rights are transferable. In these oases, the public good has not been fully transformed into a private good because the government still maintains an interest in the right and can revoke it. Someone who wishes to purchase a broadcast frequency from an existing license holder, for example, must meet certain governmentally imposed criteria or else he or she cannot take possession. The more common case is where the collectlvlty grants a nontransferable right. Such rights are granted free to wilderness users through the allocation of wilderness permits and auctioned or otherwise allocated for fees to those who wish to we public lands to mine coal or oil, to harvest trees, or to graze livestock.

Collectively held rights occur where access to the good (or potential access for unavailable quality levels) is available to all members of the collectlvlty and individual members cannot sell their access right. If there is a cost to providing the good at a given quality level, it is normally borne. by all consumers through some combination of taxes, higher prices and fees. If this level of payment is not maintained, the quality of the, good will often deteriorate. (In what follows here, we restrict our attention to goods which require recurrent payments to maintain a given quality level.) If a quality increase is desired, higher payments will be required to cover the cost of providing the new quality level. The less excludable the good, the more likely it will be collectively held since entrepreneurs cannot efficiently provide it at a profit. Water quality is a good example of a good to which individual consumers have collective, nontransferable property rights of this kind.[8] The

8. The concept of oollectlve rights must be regarded as an ideal type,
(Footnote continued)

appropriate analogy for this type of public good is not marketed goods, but maintenance fees such as those paid by condominium owners. Purchase of a condominium conveys private property rights to the apartment itself. But condominium owners are also legally obligated to pay fees, whose level is collectively determined, to maintain the property and its grounds. Owners can, if they choose, collectively agree to increase their fees in order to provide a more lavish common amenity. Nonpayment of fees or a reduction in their level would result in a lower quality common amenity. all owners have equal rights to *use" these nondivisible collective goods.

The implications of this framework for the choice of the correct Hicksian surplus measure for the present study is shown on the right hand side of table 1 for the case where the good requires recurrent payment to maintain quality. Our aim is to measure the benefits of national freshwater quality from the consumer's initial level of utility. Where a given quality level of water quality is not currently available, a CS_{WTP} measure is indicated, just as it is for measuring the consmer surplus for 8 private good which an individual neither owns nor currently uses. In both cases, the measure is the amount the consumer is willing to pay for the improvement which leaves him or her just as well off before the change as after. More unconventionally, the CS_{WTP} measure is also indicated for where 8 given, quality level is currently available. Since the consumer is already paying for this level of water quality on a regular basis through higher prices and taxes, the consumer surplus for this case is the amount the sconsumer is willing to pay to forgo the reduction in the

8. (continued)
however, since even air and water quality are transferable under certain types of administrative arrangements such thore which allow corporations to buy and sell permits to emit specified levels of pollution. The opposition to these arrangements by environmental groups is motivated in part by a belief that rights to these public goods should not be transferable.

quality level of the good and still be as well off as before. An analogy is the amount a tenant is willing to Pay in rent upon renewal' of a lease. The tenant already has possession,,but the right to retain possession at a given quality level is subject to periodic renewal. To use a referendum analogy, the consumer is asked to set the highest amount he or she would be willing to pay annually in taxes for a given program which guarantees to maintain the present level of supply of a good for the next and succeeding fiscal years. Note that the WTA format is clearly inconsistent with the nontransferable character of this property right.

Accordingly, we define the status quo in this study as the situation where, if all present annual payments were discontinued, the present level of water quality would deteriorate to below a boatable minimum quality level. This approach allows the compensating surplus-WTP measure to be used for each of the quality levels we value in this study. Respondents were asked to value each level by saying how much they are willing to pay under the (hypothetical) condition that if this amount is less than what they are currently paying they will receive a refund and that if it is higher, their taxes would be raised to cover this coat. After considerable introductory material, the value of the minimum boatabl e level was measured first by asking respondents to give:

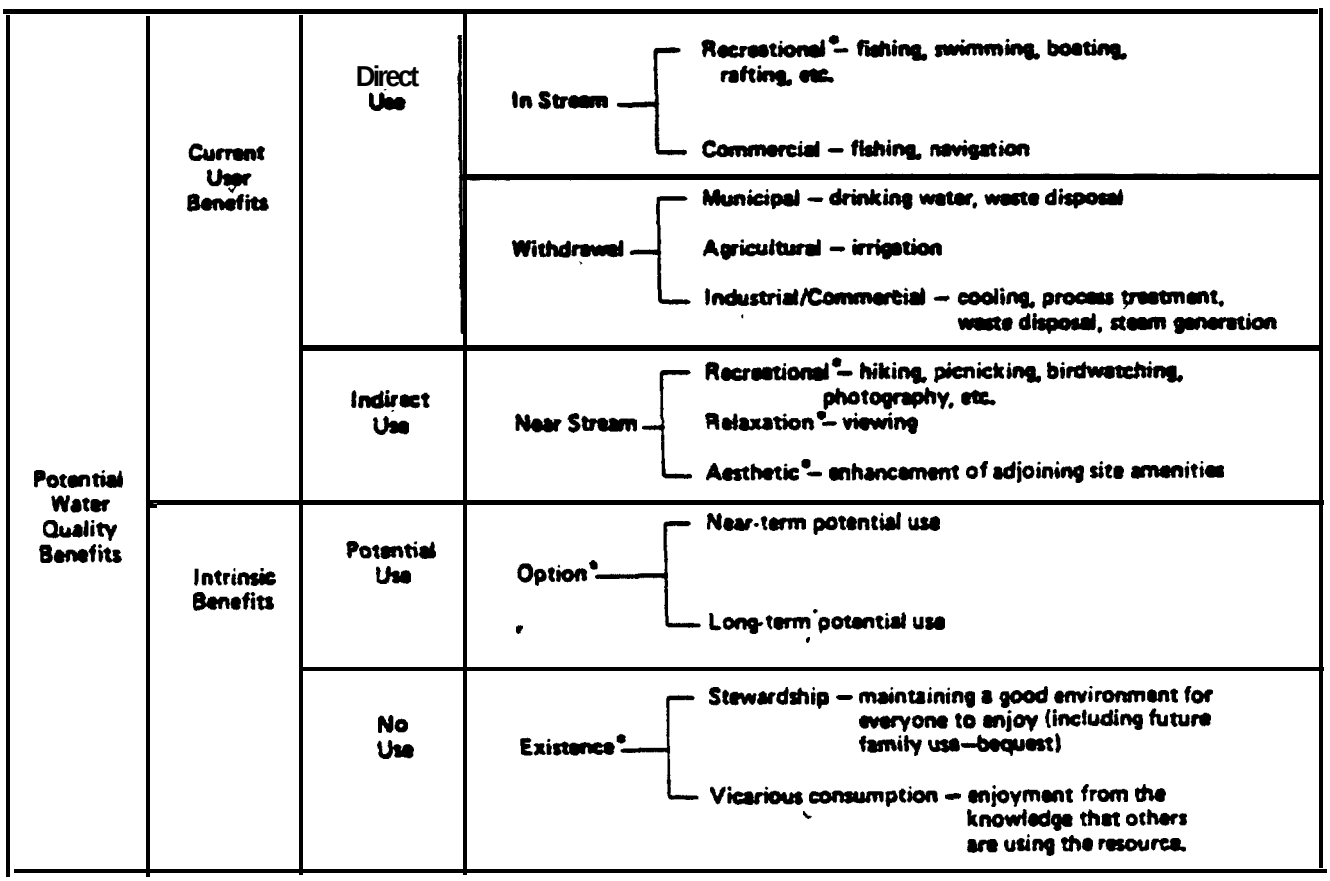
> . ..the most your household would be willing to pay in taxes and higher prices each year to continue to keep the nation's freshwater bodies from falling below the boakble level where they are now... (q. 24)

Subsequent questions asked how huch more if anything it would be worth to achieve uch of the remaining two goals, national minimum water quality levels of fishable and swimmable.

Benefits Measured

Figure 2 presents a typology of water benefits. One of the principal

Figure 3 FRESHWATER QUALITY BENEFITS



* Considered in this project.

From: Desvousges, Smith, and McGivnety (1983).

sdvankges of the contingent valuation method is its ability to measure intrinsic (nonuse) benefits in addition to recreational and indirect use benefits accruing from consumers' in and near stream activities. The benefits l asured by this study are indicated by an asterisk.[9] After pretests showed that some respondents tended to confuse drinking water benefits with freshwater quality benefits, we included a statement in the scenario warning against this interpretation. Pretests also showed that some respondents ignored the indirect benefits, so we reminded them of the full range of benefits. by adding a discussion of a "values" card (Desvousges, Smith, and McGivney, 1983) which lists the major reasons why households night value water quality. It is unlikely that the respondents took any of the commercial instream benefits or any of the withdrawal benefits into account since we told them that one of the ways in which increased water quality would be paid for is through higher prices and these benefits would likely result in lower prices for consumers.

Design Features

In this section, we describe the key features of our CV instrument.[10] Its aim is to present a hypothetical market which describes thd good and its provision in a way that is both consistent with economic theory and also understandable by as many respondents as possible. It contains various features designed to facilitate respondent understanding and to minimize the possibility of bias.

Scenario Elements

9. Much of the typology in figure 2 was first presented in Mitchell and Carson (1981) and the version here is taken directly from Desvousges, Smith, and McGivney (1983) who substantially improved it.

10. Chapter 3 describes the evolution of the instrument from an earlier version we tested In 1980 (Mitchell and Carson, 1981) and provides more details about its features. Appendix A presents the instrument l s full text.

The payment vehicle, annual taxes and higher prices, corresponds with the wsy citizens presently pay for water quality. In an effort to avoid the starting point bias associated with the bidding game method, the elicitation format uses the anchored payment card format which we developed for our 1980 study." Respondents were divided into five income groups baaed on their household income and given a payment card containing a large array of amounts, five of which-- the anchors -- were identified as the amounts average households of that income group are currently paying In taxes and higher prices for nonenvironmental public goods such as defense, the space program, and police and fire protection. The willingness-to-pay questions asked respondents to state the amount on the payment card or "any amount in between" they are willing to pay for the given water quality level. An experiment, conducted as part of our 1980 study (Mitchell and Carson,1981), tested for possible bias induced by the anchors and found the WTP amounts were insensitive to the number and dollar amounts of anchors similar to those used here. In a second experiment, conducted as part of a pretest for the present study, we compared the use of identical anchored and unanchored payment cards to see if the anchors oontribute to the quality of the data. Although the findings of this experiment are tentative owing to the small sample size (N = 93), we found no significant difference at each level of water quality between the mean and median WTPs for the two types of payment cards.[12] However, the interviewers

11. Chapters 4 and, especially 5, describe the payment card in more detail. The major changes between the payment card used in this study and our original asrd are: an Increase in the number of payment cards used (from 4 to 5) and Improved derivations of the anchor amounts using more recent data. By using an extra payment card (the original $25,000 and above upper income category was divided into a $25,000 to $49,999 and a $50,000 and above group), we were able to present much more accurate and reasonable amounts to the upper income households than before.

12. See chapter 5 for a more detailed description of these findings.

strongly felt the anchors helped respondents to arrive at meaningful answers, and it appurs that the anchors reduced the unexplainable variance in the WTP responses.

The goods to be valued -- boatable, fishable and swimmable minimum freshwater quality levels-- were described in words using the water quality ladder mentioned above as a visual aid. Compared with our 1980 study, greater emphasis wss placed in the scenario on informing respondents about the current nonuniform distribution of water quality in. the United States and the nonunif orm nature of the improvements. In order to implement our WTP-compensating surplus questions, we also added material to the scenario which reminded the respondents' that they are currently paying part of their income for the nation's water pollution control programs in taxes and higher prices. Once consequence of this material, which turned up in our pretests, was that some respondents were unwilling to answer the WTP questions without knowing how much they were paying. This created a dilemma, because Informing them of the amount could lead them to base their value on this amount instead of independently determining their maximum WTP amount. This problem was solved by telling respondents that they would be told the size of their present payment at a later point in the interview,but that it was important for them to give their WTP prior to receiving this information. This seemed to satisfy most respondents and also allowed us to test the effect of providing this information at a later stage of the interview.

Types of Pleasures

In order to provide respondents with the chance to revise their WTP amounts on the basis of a greater understanding of the valuation exercise and to measure the effect of providing additional information or incentives to change their answers, four WTP amounts were solicited from each respondent for

uch of the three water quality levels.

WTP_F The first bid is the amount given- for each of the three WTP questions (boatable, fishable and swimmable; questions 24, 26, and 27).

WTP_R The reconsidered bid is the amount (whether changed or unchanged) bid after their three first amounts were repeated to them, the total was stated and they were encouraged to make any revisions they wished (question **29).**

WTP_I The informed bid is the amount given after respondents were told how much the range of the amount households in their income group (question 33) were actually paying for water (and air) quality.

WTP_H Finally, respondents were asked if they would increase their WTP amounts if they were not enough to reach any of the three goals, Including the boatable wster quality goal. The amounts given after this question (35) is the highest bid. The results of each of these revision exercises are given below.

At this point, it is useful to make clear our assumptions about the nature of the amounts each of these bids elicits. We assume that many or most of the respondents who are asked in a CV survey to value a good which they are unaccustomed to purchase do not have a well formed value for such a good. [13] Faced with such a first-time request for such a value, some respondents are unable to offer a value. The remaining respondents, however, know within a reasonable range where their value for the good may lie and a few may even have a good idea of the actual value. n the assumption that respondents are generally cautious (i.e., risk averse consumers) when faced with sizable purchases, we believe the WTP_R amounts given by the remaining respondents are likely to represent the lower bound of their WTP range. In the case of the WTP_H amounts where the question wording implies that they hsve not given the correct number and that the “appropriate" number is higher, the WTP responses

13. Because respondents in such a CV study do not have well formed values, their WTP amounts are vulnerable to bias induced by elements of the scenario which suggest appropriate values. This means that the scenario must be designed in such a way as to minimize bias from such factors as starting points, a point which we discuss in chapter 2.

are likely to represent the range's upper bound.

A major question in valuing water quality improvements is the shape of the benefit curve between the three goals of boatable, fishable and swimmable. If people's willingness to pay is totally contingent upon the attainment of each goal, the function is a step function, and intermediate or partial improvements would provide no additional benefits. Two questions, asked of equivalent subsamples (A and B), explored respondent 's views about water policies which promise partial Improvements to one or the other of the fishable and swimmable goals. In the halfway policy question, respondents were asked (q. 30, version A) if they would still be willing to pay their revised amount for swimmable "if the best we could do was to raise the minimum only halfway from fishable to swimmable." The 95 percent question (q. 30, version B) asked respondents if they would still be willing to pay the fishable amount if "five percent of the nation's water bodies remain at the boatable level...The lakes, rivers and stream comprising this five percent would all be located in heavily industrial and/or urban locations where a lot of people live."

With a public good such as water quality which is unevenly distributed geographically, It is of interest to learn the extent to which respondents value provision of the good outside their home area'. The most reliable definition of home area for a survey such as ours was the respondents ' state. After being reminded of their total ($WTPTOT_R$) bid, respondents were asked how many dollars or what percent of this amount they would give to their state and to the rest of the nation for water improvement? In order to minimize possible strategic behavior, they were told to presume that people in other states would also divide their money honestly.

Part-Whole Bias

For CV studies which attempt to estimate aggregate benefits, three types

of part-whole bias (Mitchell and Carson, 1984a) pose potentially serious problems.[14] The first is geographical where the respondent perceives either a larger or smaller geographic area being valued than that intended by the researcher. We avoid this type of part-whole bias by directly measuring national water quality benefits, the geographic area of primary interest. State benefits are estimated by having the respondents decompose their aggregate national value. The second is benefit component part-whole bias which occurs when respondents over or underestimate their WTP for components (such as option and use) of their total value for the good because they are unable to meaningfully break their total WTP down in thfs manner. In this study we only ask for the total WTP.[15]

The third is policy package part-whole bias, where the policy package perceived by the respondent is defined more or less broadly than intended by

---

14. Part-whole bias as we describe it here is distinct from the sequencing effects noted by by Randall and his colleagues (Hoehn and Randall, 1982). These effects were empirically demonstrated in studies of regionally specific air pollution benefits (Randall, Hoehn, and Tolley, 1981). The sequencing effect occurs when the value of a particular good or policy depends on the order in which it is valued in relation to the other goods or policies in the sequence. Because goods are substitutes and complements, the sequencing effect is an understandable economic phenomenon. Aggregation bias can occur if the researcher sums up the values for each of the goods in the sequence where eaeh good was valued as if it were the first element in that sequence. Part-whole bias, on the other hand, involves a divergence between what the respondent values and what the researcher intends the respondent to value. In this study, sequencing effects are minimized by considering the entire water pollution control program for the United States.

15. Fisher and Raucher (1984) suggest a defensible means of indirectly estimating the lower bound of our sample's intrinsic (nonuse) benefits for water quality by dividing the nonusers' WTP by the total sample's WTP. When nonuse is defined as no instream recreationnal use of freshwater by the respondent In the past 12 months, intrinsic benefits calculated by this procedure amount to 39 percent of the total WTP amount. When nonuse is extended to include everyone in the respondent's household, Intrinsic benefits amount to 30 percent of the total. Finally, if nonuse is defined as no direct or indirect (e.g. picnicking, camping, duck hunting etc. by freshwater) activities by anyone in the household, a lower bound for intrinsic benefits of 19 percent is indicated.

upper bound.

water quality improvements is the shape of

ble.

goals of boatable, fishable and swimma

upon the attainment of

s totally contingent

step function, and intermediate or Partial

dditional benefits-

Two questions, asked of

B), explored respondent's views about water

improvements to one or the other of the fishable

halfway policy question, respondents were asked

ould still be tilling to pay their revised amount

e could do was to raise the minimum only halfway

version B) asked

The 95 percent question (q. 30,

fishable amount if *five

t be willing to pay the

at the boatable level.. .The lakes,

e r bodies remain

rcent would all be located in heavily

ing this five pe

ocations where a lot of people live."

distributed

such as water quality which is unevenly

which respondents

interest to learn the extent to

The most reliable

good outside their home area.

ours was the respondents' state-

for a survey such as

asked how

their total ($WTPTOT_R$) bid, respondents were

ant of this amount they would give to their state and

In order to minimize possible

for water improvement?

that people in other states would

were told to presume

onestly.

three types

h attempt to estimate aggregate benefits,

, 1984a) pose potentially serious

the respondent perceives either a

alued than that intended by the

hole bias by directly measuring

phic area of primary interest.

e respondents decompose their

efit component part-whole bias

imate their WTP for components

for the good because they are

wn in this manner. In this

, where the policy package

s broadly than fntonded by

stinct from the sequencing

oehn and Randall, 1982).

es of regionally specific

1981). The sequencing

r policy depends on the

goods or policies in the

s, the sequencing effect

bias can occur if the

the sequence where each

t sequence. Part-whole

n what the respondent

value. In this study,

ntire water pollution

means of indirectly

onuse) benefits for

sample '5 WTP. When

freshwater by the

calculated by this

. When nonuse is

intrinsic benefits

efined as no direct

. by freshwater)

rinsic benefits of

2. ESTIMATES OF NATIONAL FRESHWATER QUALITY BENEFITS

In this chapter we present estimates of national freshwater quality benefits based on the findings of the RFF contingent valuation survey. In what follows we first describe the core of usable data and the survey's findings based on these data. We then adjust these data to compensate for sample selection bias and item nonresponse bias and estimate aggregate nationai freshwater benefits. We conclude with a discussion of the costs and benefits of freshwater quality control.

Findings

Core Of Usable Responses

Of the original 813 interviews, 564 or 70 percent yielded usable WTP amounts and constitute the usable core -of interviews which we use for our data analysis. This is an acceptable item response rate given the degree of interest and effort involved in answering complex CV scenarios such as the one used in this study and represents a sharp improvement from the item response rate for our 1980 pilot study.' However, since the nonrespondents were not a random subset of the sample,it is necessary to compensate for their lack of

---

1. The item response rate for the WTP questions in that study was 50 percent. The improvement occured despite the fact that we increased the standards for l ooeptlng a WTP answer a8 valid in the present study and used a longer and more complex survey instrument. Further efforts to increase the item response rate above70 percent would face a tradeoff between increasing the usable responses **and** decreasing the proportion of the respondents who are giving meaningful answers. We speculate that the 70 peroent Item response rate may be close to the upper limit for random samples who are asked to value complex public goods. For comparison, only 13 percent more of the sample gave usable responses to a standard survey question which asked whether respondents thought "too much, about the right amount, or too little" was being spent on reducing water pollution in the nation’s freshwater bodies.

response to avoid biasing our benefit estimates. These adjustments are discussed later in this chapter after we give the initial results.

Slightly more than half of the unusable responses may be characterized as protest zeros which are zero amounts given by respondents who object to some aspect of the scenario, such a8 paying for the good by the specified vehicle, or who fail to understand the hypothetical market. These response8 were identified by a series of followup questions (qs. Y1 - Y11) asked of each respondent who gave a zero bid. If the respondents said they gave a $0 bid because. that is what the level of water quality is worth to them or because they lack enough money to pay anything, their. WTP amount was treated a8 a genuine $0. All those who gave other reasons for their zero amount were coded a8 giving protest zeros.

The remainder of those who did not respond to the WTP items, consisted of 72 don't knows (29%), 18 refusals to answer the WTP questions (7%), 16 inconsistent (too high) responses (6%), and 10 inconsistent (too low) responses (4%). Responses judged inconsistent (too high) were those which exceeded 5% of the household's Income while those judged inconsistent (too low) were WTP amounts Of less than $5.00 (usually $1.00) given by respondents with above average to high income8 and supportive positions on water pollution control expenditures. The "too low" responses may be regarded as protest zeros which messed being identified because the token positive amounts given by these respondents removed them from our protest zero screen.[2]

Bearing In mind that large differences are required when comparing percent difference8 between small samples, the data in table 1 show that people with

2. Appendix E describe8 these outliers in detail. One possible explanation of the "too high" value8 is that they represent strategic behavior. The fact that they were mostly given by respondent8 with low educational levels suggests thoughtless rather than strategic behavior, however.

the lowest level8 of education and people age 65 and over are especially unlikely to give usable responses to the WTP questions.3 Those in the oldest age category are particularly likely to give "don't know" responses.

Table 1 DISTRIBUTION OF USABLE, PROTEST ZERO AND DON'T KNOW RESPONDENTS BY EDUCATION AND AGE

Education	Grade School	Some H.S.	H.S. Grad.	Some College	College Grad.	College	Total	N
Usable WTP	6%	10	38	25	13	9	101%	529
Protest Zero8	17	20	39	16	6	2	100	135
Don't Knows	2	15	27	21	8	6	99	85

Age	Under 30	30-45	45-64	65+	Total	N
Usable WTP	27	31	24	18	100	529
Protest Zeros	15	26	33	27	101	136
Don't Knows	12	24	22	43	101	87

Unadjusted Estimates

Table 2 presents the WTP amounts for each of the four series of bids measured in the study. Using the reconsidered series of bids, the respondents who gave usable responses were willing to pay $106 annually for maintaining boatable quality water ($WTPB_R$), $80 more to reach the fishable minimum water quality level ($WTPF_R$), and an additional $69 for swimmable quality water ($WTPS_R$) for an unadjusted total ($WTPTOT_R$) of $276.

Analysis Of Changes

An examination of the changes, made by the 75 respondents who reconsidered

3. These characteristics generally predict nonusable (don't know) responses to other questions in the survey. Examination of other characteristics show these to have the strongest relationship.

Table 2. MEAN UNADJUSTED ANNUAL HOUSEHOLD WILLINGNESS TO PAY AMOUNTS FOR DIFFERENT LEVELS OF NATIONAL WATER QUALITY BY TYPE OF BID

Water Quality Level	First Bid(F)	Reconsidered Bid(R)	Informed Bid(I)	Highest Bid(H)	N
Boatable (WTPB)	$111 (10;$40)*	$106 (10;40)	$125 (11;48)	$141 (13;50)	564
Fishable (WTPF)	80 (8;30)	80 (8;30)	96 (9;35)	108 (10;50)	564
Swimmable (WTPS)	89 (12;25)	89 (12;25)	102 (12;25)	116 (13;25)	564
Total WTP (WTPTOT)	280 (25;125)	276 (25;120)	323 (27;150)	366 (29;150)	564
Number changing their bids at each stage		75	104	136	

*(Standard error of the mean; median).

and revised their amounts after the giving their WTP_F amounts, shows that many of them corrected mistakes caused by respondent misconceptions about the elici tation process. These mistakes fell Into easily recognizable patterns. One involved respondents who did not initially grasp the fact that they would be valuing three water quality levels and who therefore gave most of their water quality dollars for the first level. Given the opportunity to reconsider t hese amounts, these respondent8 typically reduced their WTP amounts for boatable quality and increased their bid8 for one or both of the other two levels. Respondent 1252, for example, initially gave WTP amount8 of $800, $0 and $0 which he revised to $100, $100 and $300 for boatable, fishable and swimmable national water quality respectively. A second pattern occured when

respondents lowered their total WTP amount after after realizing that the total they had comitted themselves to In the first sequence was higher than they intended. By far the most dramatic example of this was respondent 2268. Although her household income was only in the $22,500 category, she initially gave $300 for boatable, $400 for fishable and another $400 for swimmable water quality -- a total of $1100. Given the opportunity to revise her amount, she said she was willing to pay only $100 for each level, for a much lower $300 total. The overall effect of the increases,decreases and reallocating which took place at this stage was a very small decrease in $WTPTOT_F$.

In the WTP_I iteration, one hundred and four or 18 percent of the respondents revised their bid8 after being told the actual range of what people in their income category are currently paying in taxes and prices for water quality. Those who earlier revealed what amounted to "underpayments" relative to what they are paying were more likely to change their WTP amounts to bring them closer to the actual payments than those who "overpaid." This behavior is consistent with a range of plausible motivations ranging from conformity to the socially approved goal of paying your fair share to rethinking the value of water quality in the light of information thought to reveal something about its true price. Approximately half of the changers were respondents who increased their WTPTOT upon discovering that their WTP_R amount was below what they were paying whereas only fourteen lowered their $WTPTOT_R$ amounts upon discovering these amount8 were higher than their current payments. Of the other changes, increases predominated including seven people who increased their $WTPTOT_R$ amounts from the actual payment range to a level higher than this range. The overall effect of these changes on the total WTP amount was a substantial 17 percent increase spread quite evenly across the three water quality levels.

The last Iteration, where the respondents were confronted with the assertion that the amount they had previously committed themselves to might not be enough to maintain even the present minimum level of water quality (a strong statement), stimulated the largest number of change8 and produced a further 13 percent increase in WTPTOT. One out of four respondents increased[4] their amount8 at this stage including 34 people who already had given amounts which exceeded their actual payments. Overall, taking those who made multiple changes into account, approximately 30% of the respondents changed one or more of their WTP amount8 and of these about a third changed more than once.[5]

WC believe the WTP_R series represents the most valid basis for estimating WTP (after adjustment for nonresponse) because the informed and, in particular, the highest series of WTP questions put a significant amount of social pressure on the respondents to revalue their responses upward. The prospect offered respondents in the highest condition was quite drastic -- that even the boatable level was threatenedif a higher WTPB bid was not forthcoming. We will show the usefulness of the two post-WTP_R series later after wc first turn to the issue of explaining why people gave the WTP_R responses they did.

Estimation

Using a theoretically based equation to predict WTP amounts in CV surveys provides evidence for the reliability of the data as well as for its (construct) validity. Based on a reading of the relevant literatures and our earlier work (Mitchell and Carson, 1981) we hypothesize:

$$WTP = f(INC, WSPEND, USE, ENV, GOVT) \quad (1)$$

---

4. No one decreased their WTP amounts at this stage.

5. The mean bids for the first and reconsidered condition8 were not significantly different except for WTPB. Each of the other two revision opportunities gcncraliy resulted in mean bids which were significantly different from their predecessors.

where INC is household Income, WSPEND is the respondent's attitudes toward water quality expenditures , USE is the household's recreational use of water, ENV Is the respondent's self identification as an environmentalist and attitudes toward general environmental Issues,and GOVT is the respondent's attitude toward government spending and regulation and the respondent's beliefs about the government's efficacy in providing public goods. Operationalization of INC is straightforward (q. 40, in thousands). WSPEND is measured by constructing a five point scale from questions lo, 2, and 3, giving a score of 5 to "spend a great deal more" on water quality. We gathered a large amount of household recreation data in the questionnaire. The USE measure adopted here is total person days of water-based recreation by household members in the past year, TDUSE. Although we included questions on pollution and pollution control cost in our 1983 instrument,' we were unable, for reasons of length restrictions, to include any questionswhich measured the respondent's self identification as environmentalist. Nor did we include any questions which-pertained to GOVT.[7]

Using a simple linear functional form and the available variables, we have as an estimate of equation (1):

$$WTPTOT_R = \underset{(-4.6)}{-347} + \underset{(11.1)}{8.75} \bullet INC + \underset{(5.2)}{99.26} \bullet USPEND + \underset{(2.4)}{.51} + TDUSE \quad (2)$$

where the numbers in parentheses are the t-statistics and the adjusted R^2 is

6. The responses to these questions are highly correlated with WSPEND.

7. In retrospect (and based upon our subsequent analysis of the Mitchell and Carson [1981] survey data which included all of the questions used on the 1980 Council on Environmental Quality survey [Mitchell, 1980]), the inclusion of questions related to ENV and GOVT would have been useful and should be included in future CV surveys. On the other hand, our data analysis for this survey finds that much of the detailed household recreational use information obtained is redundant for the purpose of predicting WTP.

.26.[8] Looking at the residuals snd the statistics based upon them such as Cook's D and Belsley, Welch, and Kuh's DFFITS and DFBETAS, it appears that the poorest predictions and most influential cases involve respondents with large *incomes* who are willing to pay large amounts for water quality improvements.

The log-log form of equation (2) rectifies much of this problem:

$$\underset{}{\text{LWTPTGTR}} = \underset{(-1.5)}{-.52} + \underset{(12.9)}{1.03} \bullet \text{ZINC} + \underset{(7.0)}{1.54} \bullet \text{LWSPEND} + \underset{(2.0)}{.06} \bullet \text{LTDUSE} \qquad (3)$$

where the L prefix indicates that the natural log has been taken.[9] This equation gives an adjusted R^2 of .36 which indicates a good fit given cross-sectional data and such a parsimonious model. An examination of the diagnostic statistics for this equation shows a much better overall fit. Any problems which now exist tend to lie with respondents who have moderate to high incomes and very low WTP (typically zero or one dollar). Since we suspect that some of our zero or low token amounts are really protest zeros which were not picked up by our protest zero screen, this is a more intuitive pattern of problem responses.[10]

Part-Whole Experiment

The results of our test for part-whole bias are given in Table 3. Recall that respondents in subsample A were informed of what they were currently paying for water quality whereas those in subsample B wre told the amounts for

---

8. The regression is based on 481 observations. The difference between our 564 "useable" responses and the 481 observations used in the regression equation is due to missing values on the independent variables particularly INC and WSPEND.

9. Since the log of zero is undefined, the following two conventions were used. LTDUSE = 0 if TDUSE = 0 and Log(TDUSE + 1.7128) if TDUSE > 0. The minimum $WTPTOT_R$ was set to 1 (i.e., $WTPTOT_R$ < 3 set equal to EXP).

10. Based upon our previous experience, we suspect that many of the WTP amounts given by these people would have been predictable if a GOVT variable were available.

both the water and air pollution control programs. (For example, B subsample households in the $20,000 to $29,999 income range were told that their present payments for water pollution are $175.$300 per household and their air payments are $265-420 whereas the comparable respondents in subsample A were only Informed of their water pollution payments.) If part-whole bias is present we would expect less of a positive difference (D) (or a greater negative difference) between the reconsidered bids and the informed bids for treatment B because of a greater propensity on the part of those respondents to correct overspending on water pollution control by reducing their bids. The Z value for the non-parametric Wicoxon test statistic in each case is very insignificant [11] and the hypothesis is rejected. [12]

Table 3. TEST OF INTRODUCING AIR POLLUTION CONTROL COSTS

	Mean(A)* (SEM A)	Mean(B) (SEM B)	Z(Wilcoxon)	Prob>|Z|
DB** (Boatable)	**$21.85 ($7.03)**	**$14.57 ($6.63)**	**0.35**	**.73**
DF (Fishable)	**12.87 (5.97)**	**14.79 (7.50)**	**-0.53**	**.60**
DS (Swimmable)	**10.41 (4.82)**	**7.44 (4.73)**	**0.51**	**.61**
DTOT (Total)	**45.14 (12.37)**	**36.80 (15.65)**	**0.84**	**.40**
	N=311	**N=253**		

*Version A informs respondents of their present payments in taxes and higher prices for water pollution control. Version B informs them of their payments for both air and water pollution control.

**DB = Informed WTPB - Revised WTPB.

In and Out-of-State Benefits

Our data show that many respondents have some value for out-of-state provision of the good. Although most people (more than 90 percent) wanted at least one third of their $WTPTOT_R$ spent in-state, only one person out of three wanted all of It spent in-state. Overall, the respondents answering the WTP questions allocated 67% of their $WTPTOT_R$ for water quality improvements to be spent in their state and 33% of this amount to be spent out-of-state.[13] The median in-state percent (70%) was almost identical to the mean. The correlates of preference for local benefits suggest that the more cosmopolitan a person's orientation and experience, the greater their allocation for out-of-state benefits. In-state benefits wre positively correlated with the number of years lived in the state and age and negatively correlated with education, income (to a lesssr degree) and recreational use of. out-of-state water. There were no significant differences in the in-state/out-of-state split across regions.

Partial Improvements

According to the answers to the'halfway" and "95 percent" questions, the

---

11. The apparent large size of the difference between versions A and B is an artifact of the method used to test the hypothesis. Because the pre-Informed WTP amounts for A and B are not quite equivalent, we compare the incremental change from the revised to the informed WTP measures for each version instead of the absolute WTP_I amounts. The $7.28 difference between the A and B increments for boatable water quality is less than 6 percent of the WTP amount for this quality level, and well within the confidence interval suggested by sampling variation.

12. We also performed t-tests on the actual differences and the difference of the logs. The largest t-value was .74. In these cases the t-test is grossly Inefficient because, since most respondents did not revalue, the distribution resembles an extreme version of a double exponential with a large spike at zero. me Wilcoxon test is much **more** EFFICIENT in this case and only slightly less efficient in the normal case (Lehmann, **1975**).

13. This estimate had a standard error of the mean of 1.18 percent and was based on 530 observations.

benefits of partial Improvements are considerable. Almost nine out of ten (89 percent) of those who answered the question said the 95 percent improvement from boatable to fishable is worth the same to them as the complete improvement.[14] Those who were unwilling to pay the same amount for the partial improvement in this case were disproportionally residents of large urban areas. This is understandable because the question informed respondents that the "lakes, rivers and streams comprising this five percent would all be located in heavily industrial and/or urban locations are a lot of people live ." Each person who was unwilling to pay the same amount was asked how much they were willing to pay for this partial improvement. The WTP amount for raising 95 percent of the nation's water to at least the fishable level is $74 or 8 percent less than the $WTPF_R$ for raising 99 percent to at least that level. Turning now to the halfway improvement question, which was asked of subsample A, we find a somewhat lesser percent (73 percent) were willing to pay the same amount for the halfway improvement from fishable to swimmable as they were for the total improvement. Because those who were not willing to pay the same amount wre willing to pay a somewhat greater percent for the partial improvement than In the 95 percent case, the overall reduction in $WTPS_R$ for swimmable water quality is slightly less.

It is possible to compare these estimates of the benefits of the 95 to 99 percent fishable water partial Improvement with a recent estimate made by Vaughan and Russell (1982) using a participation-travel-cost model. Vaughan and Russell valued the benefits accruing to fishermen from improving national freshwater so that all waterbodies are at least at the fishable quality level. This improvement is equivalent to raising three to five percent of the

14. Which w defined as where 99 percent or virtually all the nation's lakes, streams and rivers would be fishable.

waterbodies from less than fishable to fishable quality, an increase quite similar to the 95 vs. 99 percent improvement w asked our respondents to value. It might be expected that our estimates should be somewhat higher than Vaughan and Russell 's due to the more Inclusive nature of our benefits. On the other hand, it is likely that recreational benefits dominate the benefits of the 95 vs 99 percent improvement in our survey, since the 95 percent level provides for a large number of available substitutes and is likely to fulfill many people's stewardship needs. Vaughan and Russell's estimate of the benefits for this improvement range from 200-1200 million (1983) dollars with 500 million dollars as the best rough point estimate. Considering the difference in methods and data bases, this amount is remarkably similar to our 490 million dollar point estimate for the 95 to 99 percent improvement.

Distribution of Benefits

Baumol and Oates (1979) have noted that studies of the distributive effects of environmental policy are still in their infancy despite the crucial importance of the equity issue for environmental policy. Based on their review of the then available literature on distributive benefits, they raise the possibility that the less affluent may believe that environmental improvements come at their expense. They cite poll data as evidence for a "consistent pattern of disproportionately strong support for environmental programs among higher-income groups" (Baumol and Oates 1979: 184). One of the major advantages of the CV method over other benefit estimation techniques is the Information It provides on the distribution of the benefits for the program being valued, thus permitting the identification of losers and the gainers.

The data presented in table 4 enable us to assess the distribution of water quality benefits for five broad income categories. It shows, first, average willingness to pay for water quality Increases sharply with income. In

absolute terms, the respondents in the highest income category are willing to pay almost 19 times as much, on the average, as those in the lowest income category. However, when benefits are measured by the percent of their income people are tilling to pay for water quality, the distribution is only mildly progressive. The biggest difference is between the lowest and highest income groups with the middle three groups showing no clear pattern. This finding is quite consistent with the pattern of broad-based pattern of support for environmental goals snd the environmental movement which became apparent in numerous public opinion polls in the late 1970s (Mitchell, 1979; and the 1980s (Council on Environmental Quality, 1980; Ladd 1982; Mitchell, 1984), and in the distribution of responses presented in table 4 to a question in our present survey which asked respondents how important to them personally is a "national goal of protecting nature and controlling pollution." As indicated there, at least 60 percent of every income group said such a goal is "very important" to them personally with only modest (and Insignificant) differences between the high and low income groups.

It thus appears that demand for environmental quality in general and for improved water quality in particular is broad based although the monetary benefits are subject to strong income constraints. Two recent studies of the distribution of water pollution control costs (Lake et al., 1979; Gianeesi and Peskin, 1980) found that they tend to be mildly regressive overall and especially regressive at the lower income levels, [15] because these costs are

15. Comparisons between the two studies are made difficult because of differences in their baselines and demographic projections. However, both show the lowest income group is paying more than twice the percent of their Income toward water pollution control than those in the highest income groups. The regressive impact of water pollution control costs is mitigated somewhat by the federal sewage treatment plant and nonpoint source control programs. Control costs for air pollution are more regressive (Gianessi and Peskin, 1980) owing to the absence of comparable federal programs.

Table 4. WTP_B FOR WATER QUALITY UP TO SWIMMABLE LEVEL WATER AND IMPORTANCE Of NATIONAL GOAL OF PROTECTING NATURE AND CONTROLLING POLLUTION BY INCOME GROUP

Household Income	N	Mean	Std. Error	Median	As % of Income	National Goal of Protecting Nature and ControllingPollution Very Important*
under 10,000	125	$ 61	$ 6	$ 35	.90	60%
10,000-19,999	154	171	16	100	1.18	71
20,000-29,999	130	225	20	150	.92	66
30,000-49,000	97	422	45	270	1.13	63
50,000 and over	41	1154	281	600	1.32	66
All Respondents**	564	276	25	120	1.05	66

* Question wording: "Some national goals are more important to people than others. How important to you personally is a national goal of protecting nature and controlling pollution? Is it very important, somewhat important, or not very important to you."

**Including those who did not give their household's income.

paid largely through sewer fees and higher prices for a number of basic consumer goods. Present water pollution control policies sre therefore inequitable, a finding made more apparent by treating our CV survey as an analogue of a voter referendum. If the referendum was on a flat tax, the median votsr would rule and $120 is the maximum annual amount that would be approved by a majority. If the referendum proposal was for a progressive tax, with each of our broad income groups paying the median amount for that group, the Indicated overall average payment is $164. Roth of these amounts are far short of our sample mean of $276, although it should be noted that our income

categories are fewr than the income brackets on which differential tax rates are based and therefore may underestimate the amount that would be-approved by an ideal referendum on a progressive tax. Greater equity will be achieved if a larger portion of the costs of water pollution control are collected by Income taxes.[16] The fact that water users account for a disproportionate number of those in each income group who are tilling to pay more than their group's median WTP amount, suggests that full equity would require increasing the amount collected for water pollution control by means of recreational user fees.[17]

We can also examine the regional distribution of water quality benefits. The design of our sample gives us comparable subsamples for the four census regions shown in table 5. The medians provide the best inter-regional comparisons because the mean estimates from small subsamples are prone to bias introduced by a few unusally large amounts. Although the regions do not differ significantly on the means, some of the median differences, especially between the Northeast and the South, are highly significant. These regions are broadly defined, however; a larger sample designed to provide sufficiently large random subsamples for each of the nine census regions might show differences.[18]

Table 5. $WTPTOT_R$ BY CENSUS REGION

Region	N	Mean	Std. Error	Median
Northeast	115	$256	$30	$160
Midwest	151	281	44	110
South	174	283	40	100
West	124	231	31	125

Aggregate Estimates

Adjusting for Sample Selection and Item Nonresponse

The WTP_R (reconsidered) series wre adjusted for item nonresponse and sample selection bias in two phases.[19] In the first, which imputed WTP values for the thirty percent of the respondents with missing WTP values, we assigned each observation to one of six categories ordered by $WTPTOT_R$[20] and used CART, a tree structured classification procedure recently developed by Breiman et al. (1984), to estimate a classification tree. This tree is given in figure 3. The square boxes in the tree represent terminal nodes and were used as the imputation olasses. Each observation with a missing/unusable value for $WTPTOT_R$ was classified into one of the terminal nodes. The missing WTPTOTR values for these observations wre imputed by randomly assigning values to these observations taken from that node's pool of valid WTPTOTR values.

While CART is a very powerful non-parametric technique which has much to recommend it in situations where economists are currently using logit or probit, the feature which is crucial for our purposes is Its surrogate splits. These identify the alternate splits which can be wed in place of the optimal split. For example, the first split in the CART tree in figure 3 shows that

---

16. Nonpoint source controls and subsidized sewage treatment plants are the primary diect Federal expenditures on water pollution control.

17. The upper Income groups may also view part of their income transfer to the lowr income groups as earmarked for payments for sewage treatment.

18. For instance, the West category would be split into the Pacific and Mountain census regions. According to the findings of our 1981 study, the Pacific Region has a substantially higher willingness to pay for water quality than the Mountain Region.

19. The procedures wed in this section as well as a number of alternative methods are described in Carson (1984).

20. The categories and their labels are: 1: $WTPTOT_R$ $0-$25; 2: 26-74; 3: 75-149; 4: 150-249; 5: 250-499; 6: 500+.

households with an income of less than or equal to $15,000 go down the tree to the left and those with a greater income go to the right. What if, as is the case With our data, a large number of the respondents Who failed to answer the WTP questions also did not answer the Income question? CART solves this problem by estimating the splits on the other variables which best mimic the optimal income split and wing these splits for the observations for which data on the optimal variable are missing. In our example, age is the best surrogate split variable and observations With a missing income value are accordingly sent left or right on the basis of age.[21]

In the second phase, we used the household weights supplied by the Opinion Research Corporation to weight the observations to make the sample representative of the Census population. As is typical in national probability sample surveys, women were somewhat overrepresented in our unweighted sample of respondents and young black males wre underrepresented.

Table 6 presents our adjusted estimates. A combination of household weights and imputing the missing values reduced the adjusted $WTPTOT_R$ value by 12 percent with each of the two correction techniques contributing approximately equally to this reduction.[22] Two more common methods of imputing missing values, using the mean values based on"hot deck" imputation classes developed from combinations of 'the demographic variables and maximum likelihood imputation, resulted in very similar values in the adjusted $WTPTOT_R$ -- $246 and $237 respectively. Thus, if the mean value is the primary concern, the choice

21. CART also provides useful information about the structure of the public's willingness-to-pay. hus, although the tree in figure 3 is in general agreement with our regression results,it suggests complexities which otherwise Would not be apparent and Would be difficult to model In a regression framework.

22. This scale factor was applied for consistency to the rest of the WTP_R series as shown in table 6.

of how to Impute the missing values is not critical. However, the method of using an ad hoc combination of demographic variables does not use all of the available information in the data set and the EM maximum likelihood procedure is very sensitive to the normality assumption.. The non-parametric CART procedure avoids both of these problems and provides an informative picture of the problem's structure.

Table 6. ADJUSTED ANNUAL HOUSEHOLD VALUES FOR BEST ESTIMATE OF NATIONAL WATER QUALITY BENEFITS*

	Mean	Standard Error of the Mean	95% Confidence Interval
WTP_R (Boatable)	$93	$8	$77-109
WTP_R (Fishable)	70	6	58-82
WTP_R (Swimmable)	78	9	60-96
$WTPTOT_R$	242	19	205-279

Stability of WTP Values

A comparison of the results obtained by our 1980 pilot study with the relevant WTP amounts from the 1983 survey provide an indication of the stability of these findings. Although the 1980 survey wed a shorter and less refined scenario,and had a 20 percent higher nonresponse rate to the WTP questions, it is comparable with the 1983 scenario in all important respects including the levels of water quality valued, the elicitation method, the payment vehicle and the we of personal interviews with a national probability sample. The 1980 survey produced an uncorrected estimate of WTP for swimmable

quality water of $225. After adjusting this estimate[23] to make it comparable with our 1983 estimate, a revised estimate of $252 is obtained. This amount is not significantly different from the 1983 estimate of $242.24 This indication of reliability is reassuring as we would not expect to find changes in question wording for noncentral features of the scenario causing large differences in WTP.[25] The stability in the WTP amounts in the two surveys also mirrors the stability in public attitudes toward water quality and pollution control expenditures during this time period (Mitchell, 1984).[26]

Aggregate Benefits

We can now assess the aggregate benefits implied by our data and compare them with the present costs of water pollution control. In making these comparisons, it should be noted that the benefits measured by our instrument do not include withdrawal benefits nor the benefits of preventing possible long term damage caused by the deposition of toxic chemicals in waterways and lakes.

---

23. Three corrections had to be made. (1) Since the 1980 WTP amounts were truncated at $999, the 1980 amount was ad justed by calculating the 1983 WTP with and without a $999 truncation and adding the resulting amount ($41, after discounting) to the 1980 mean WTP. (2) A multiplier of 1.20, based on the consumer price index, was wed to adjust for inflation. (3) The differential nonresponse rates were adjusted for by using a deflator of .79 obtained by the CART technique.

24. A Priori, we might have expected the 1980 WTP to be somewhat higher than the 1983 WTP since • dxoniahments against valuing other environmental quality changes were stronger in the latter survey. The WTP for boatable and fishable water quality are not directly comparable due to the significant number of respondents in the 1980 survey who apparently did not realize they would have a chance to value levels of water quality higher than boatable. Hoehn and Randall (1982) show why the Intermediate steps but not the total is affected by this type of behavior.

25. To the extent that the two surveys are different, the similarity in results is also evidence of convergent validity.

26. In contrast, WTP amounts for control of toxic waste dumps would not be expected to be stable, because public concern about this good is recent and relatively volatile.

Commerce Department estimates (Faber, Dreiling and Rutledge, 1984), put water pollution control expenditures in 1982 at $22.2 billion and pro ject them to be approximately the same in 1983.27 According to the most recent estimates by the EPA (Environmental Protection Agency, 1984), substantially higher annual expenditures will be necessary during the remainder of the decade in order to implement the best available water pollution control technology (BAT) which will be necessary to achieve the uniform goal of swimmable quality water called for by the Clean Water Act.

If we take our adjusted $WTPTOT_R$ value to be an estimate of the lower bound for household willingness to pay to achieve a water quality goal of 99 percent swimmable water, an aggregate national benefit of $20.3 billion[28] is indicated for possible benefits. The 95 percent confidence interval for this estimate is $17.0 - $23.5 billion. $WTPTOT_H$ provides a possible upper bound for these benefits at $30.7 billion with a 95 percent confidence interval of $25.9 - $35.5 billion. Using the entire spread, we would have benefits for swimmable water of between $17.0 and $35.5 billion.

As always,such numbers need caveats and qualifications. We believe

---

27. The national water quality benefits which Freeman (1982) estimated on the basis of his review of the then existing studies are not directly comparable to ours. In the first place, his estimates are baaed on the Council of Environmental Quality's (1979: pp . 666-667) definition of incremental pollution attainment due to current Federal regulation (Freeman, 1982: 5). Our benefits, in contrast, are comparable with the Commerce Department's definition of total pollution control ooat.Second, he includes marine benefits (accruing to freshwater improvements), commercial and withdrawal benefits which we do not directly measure. Freeman finds these benefits account for slightly less than half of the total water benefits and they occur primarily at the fishable or below levels of water quality. According to both Freeman (1982) and Feenberg and Mills (1980), the greatest uncertainty lies in the size of the recreational and Intrinsic water benefits with which this study is concerned. We should also note that neither we nor Freeman take account of the possibility of very long term damage due to toxic chemical contamination.

28. Baaed on 83,918,000 1983 households (U.S. Census Bureau, 1984).

$WTPTOT_R$ to be a fairly reliable lower bound. At each juncture where we were faced with a design decision which had the potential for biasing the WTP responses up or down, we chose the procedure which had the latter tendency. Virtually every one of our atatiatical adjustments for sample selection and item nonresponse have further reduced $WTPTOT_R$. [29] On the upper bound aide, we feel leas confident. lo atatiatical adjustments have been made and the respondents were subjected to possible implied pressure to raise their bids. An adjusted $WTPTOT_I$ (informed) estimate would provide a more reasonable basis for an upper bound. The aggregate $WTPTOT_I$ value is $24.0 billion. Taking this as an upper bound and aggregate $WTPTOT_R$ as the lower bound, a range of **roughly** $20 to $24 billion is indicated for the water quality benefits valued in this study.

29. Among the design decisions which potentially biased the WTP amounts downward are :presenting material about other costly public programs (e.g., crime) and tradeoffs between environmental quality and cost at the beginning of the questionnaire; reminding respondents that money spent on water pollution oontrol will **not** affect air pollution; using a payment vehicle of higher prices and taxes; using an annual rather than a monthly payment vehicle, and emphasizing to respondents that they would obtain a tax reduction if the value they give for water pollution control is leas than they are 'currently paying. The effect of the statistical adjustments -- removing outliera and adjusting for sample bias and item nonresponse -- significantly lowered the mean unadjusted WTP amounts.

3. ISSUES OF RELIABILITY ANDVALIDITY

Introduction

In this Chapter we consider the question of the quality of our data. To what extent are they valid and reliable? Are we actually measuring consumer's willingness to pay for freshwater quality and, if 80, how accurate is our measure? If there were an agreed upon criterion against which these data could be compared, our task would be straightforward. Unfortunately, the nature of public goods is such that suitable criteria are almost always unavailable and this case is no exception. Therefore, a more complex and judgmental program of assessment is called for. It involves building a plausible case that our data are not biased by the most likely sources of error. The evidence is part qualitative, part quantitative.

What are the most probable threats to a CV study's reliability and validity? Given the numerous sources of possible error in survey research, a theory-based error framework is needed to identify the most important sources and the conditions under which they pose a threat. There has been some discussion of this topic in the CV literature and a series of important mithodological experiment8 have been conducted to test for the presence of several type8 of bias, such as starting point, strategic and hypothetical. The framework we we in this ohapter is the result of our efforts to rethink these sources of error and to relate them to the concept8 of reliability and validity. It includes a typology of the most important potential biases in CV studies whioh is based on the existing CV literature and on other relevant

sources of theory from such field8 a8 survey research and social psychology.

Sampling bias in survey8 is the subject of a considerable literature (e.g. Cochran , 1977 and Sudman, 1976) and Its properties are well recognized. Somewhat more recently, the nature of nonsampling error in social surveys has received systematic attention (Dijkstra and van der Zouwen, 1982; Rossi, Wright and Anderson, 1983) as evidenced by a growing number of methodological experiments (Schuman and Presser, 1981). Because of their need to measure attitude strength with a much greater precision compared with other types of surveys, CV surveys face a particularly difficult measurement task. For example, it is not sufficient for a CV study to simply measure, as do ordinary attitude surveys, whether people are willing to pay a “great deal," a "fair amount" or "only a little" for "better" water quality. Instead, much more detailed information is required in the form of the highest dollar amount people are willing to pay over a specific time period for a specified water quality Improvement in a given location. This requirement, as we shall see, can itself promote error because of Its demands on the respondent.

Errors in CV surveys fall into two general classes: Those causing bias In the estimates, and those increasing the variance of the estimates. Presuming that the questionnaire would otherwise measure the correct phenomenon, the former affect8 validity, or whether the study is measuring what it is intended to measure, and the latter affects reliability, or the consistency of the responses. This division is not absolute as there are survey features that lessen both bias and variance, those which contribute to both, and those which pose a tradeoff between the two.

In what follows, we begin with a lengthy section which reviews the most Important potential biases which can occur in CV surveys and the techniques which we and others have used or could use to detect or minimize these biases.

We next consider the factors which affect the variance of CV estimates, Including those posing a tradeoff between bias and variance. In the final section we discuss the empirical evidence for our data's reliability and validity. Overall, we believe the design feature8 of the instrument and the available empirical indicators of reliability and validity support the conclusion that this report'8 water benefits estimates are meaningful and reasonably free from bias.

Bias in Willingness to Pay Amounts

Figure 1 present8 a typology of potential biases in CV surveys which we will use to examine the potential for bias in this study. Hypothetical bias is frequently listed in the standard litany of potential biases. On the basis of the approach adopted here, however, we conclude that "hypothetical bias" is a misnomer since there is no one bias which uniquely result8 from the hypothetical character of CV surveys. The hypothetical character of a CV study may make it vulnerable to one or more biases an/or it may affect the reliability of it8 findings. For example, some respondents, when placed in a situation where they are very uncertain about the value they hold for a good, are tempted to rely on one or more aspects of the scenario for clues a8 to the good's "correct" value, instead of making the effort to determine the value they hold for the good. Alternatively, unreliability, may occur if uncertain respondent8 answer the questions by making "wild quesses."[1] The biases which appeared to pose the most difficult problems in this study are starting point bias, budget and amenity misspecification, and item nonresponse bias. A great deal of our effort was devoted to developing way8 to minimize bias from these sources or, in the case of item-nonresponse bias, to compensate for bias from

1. Further discussion of many of these issues may be found in Mitchell and Carson (1984).

this source.

Figure 1: A Typolog of Possible CV Biases

I. Incentives to Misrepresent Responses

A. Strategic Bias
B. Comliance Bias
   1. Sponsor Bias
   2. Interviewer Bias

II. Multiple Valuation

A. Vehicle Bias
B. Method of Provision Bias

III. Implied Value Cues

A. Starting Point Bias
B. Range Restriction Bias
C. Yea-Saying Bias
D. Relational Bias

IV. Misspecification of Market Scenario

A. Vehicle Misspecification
B. Budget Constraint Misspecification
c. Amenity Misspecification
D. Probability of Provision Misspecification
E. Context Misspecification

V. Aggregation Bias

A. Sampling Design Bias
B. Nonresponse Bias
C. Item nonresponse Bias
D. Sequence Bias

Incentives to Misrepresent Responses

The first major category of potential biases in CV surveys result8 from incentives to respondents to misrepresent their stated WTP amount. Ever since Samuelson's seminal 1954 article on the nature of public goods, economists have generally held the view that people will lie when asked about their preferences

on surveys and a large number of papers have been writen on possible technique8 to get around this problem (Clarke, 1980). According to this view, there is a great danger that respondents giving WTP amounts in CV survey8 will engage in deliberate strategic behavior in an attempt to Influence either the future payment and/or provision of the public good in question. Public opinion researchers have held almost the opposite view;;that people are motivated to tell the truth, but are prone to shape their answers to please either the Interviewer or sponsor,especially when they do not have a strong or well considered view on the topic (Schuman and Presser, 1981). In this case, one would expect respondents to shape their WTP amounts in an effort to please (comply) with the perceived wishes of either the sponsor of the survey or the Interviewer.

Strategic Behavior Until recently, economists have tended to ignore the threats to validity posed by other types of nonsampling error, or they have considered these error8 secondary to the strategic bias problem. There is considerable evidence, however, which shows strategic behavior occurs with far less frequency than economic theory would predict and that it need not pose an insuperable obstacle to measuring WTP In most CV surveys.[2] Successful strategic behavior requires knowledge of the relevant parameters of the survey (e.g., mean, variance, and number of respondents) which are generally unavailable to the ordinary respondents. Indeed, the only person clearly found to be engaging in strategic behavior in one test for It in a CV study was an economics professor at a junior college interviewed at random (Rowe et. al., 1980) and a class of economic8 graduate students was by far the group with the highest level of strategic responses In a series of related studies conducted

2. This evidence is reviewed in Mitchell and Carson (1984).

in a variety of settings (Marwell and Ames, 1981). The few cases where it has been shown to occur are situations where respondents feared the immediate imposition of a user fee for a good which was provided free (e.g., Seller, Stoll and Chavas 1983). In the case of the present study, the eventual imposition of higher prices and taxes for water quality is unlikely to trigger strategic behavior and no evidence for it was discovered in our pretesting.

Sponsor and Interviewer Bias Overall, respondents who agree to participate in surveys are remarkably cooperative. They are motivated, sometimes strongly, to meet the expectations of the interviewer. Although this motivation make8 survey research feasible, it has the potential to promote bias as respondent8 may shape their answers to comply with what they take to be the desires of the organization on whose behalf the survey is being conducted or the perceived expectations of the interviewer. Often the Identity of the sponsor and/or the particular purpose of a survey is deliberately kept vague If bias from this source is anticipated. Avoidance of interviewer bias is promoted by rigorous training of interviewers which includes the inculcation of a strict rule not to deviate from the text of the interview. If a respondent has difficulty understanding a question or asks for more information, standard survey research practice require8 the interviewer to either repeat the question as written or,If the relevant material is provided, to offer a standard predetermined response to the question.

We endeavored to minimize sponsor bias by wing an introductory statement which only conveyed general information about the study's purposes -- that we wanted to know how much public programs are worth to the respondent and that their views will be used to help policy maker8 make informed decisions. The interviewers identified themselves as Opinion Research Corporatlcn employees. Only if a respondent specifically asked who was the study's sponsor were they

told that it was the Environmental Protection Agency. Few make such a request. It is impossible to test for interviewer bias if the Interviewers are assigned to geographical areas, a8 they were inthis study, instead of randomly to respondents. Any difference in answers between interviews conducted by interviewer A, whose assigned area is a lower income urban location, and those conducted by Interviewer B, who conducted all her interview8 In a suburb, will normally be attributed to population differences. Unless interviewers conduct large numbers of interviews, population effect8 cannot be disentangled from interviewer effects. The interviewers used in this study were all experienced professionals. Since more than 100 worked on the survey, the potential for bias from an Individual Interviewer is relatively low.

Multiple Valuation

Multiple valuation occur8 when the respondent simultaneously values both the specified good and another closely connected good. One form of this is bias caused by the choice of payment vehicle. The ideal payment vehicle is one which is plausible and value-neutral. Typical payment vehicles used in CV survey8 are park entrance fees, increases In utility bills, property taxes, sales taxes or Income taxes. A number of studies show that the public's willingness to pay for public goods issfrequently influenced by type of payment vehicle, contrary to the expectations of economic theory (Rowe et al., 1981; Greenley, Walsh and Young, 1981; Brookshire, Randall and Stoll, 1980). The recent practice among CV practitioner8, Including ourselves in this study, has been to use the relatively neutral vehicle, "higher taxes and prices" Since polls show consumers have negative feeling8 about taxes, use of this vehicle is likely to induce the respondent to treat the valuation process seriously and to keep in mind his or her budget constraints.

The particular method of provision (or the agent providing the public

good) wed in a CV scenario oan also bias the WTP amounts. Provision by public charities auoh a8 the Salvation Army tend to evoke higher WTP as do some agencies such as state fish and game commissionns. Conversely, provision by the Federal government ● vokua skepticism in some respondents about the likelihood that the good will be provided, even if paid for, and/or the view that the Federal government generally wastes money and should not be given additional money under any circummstance. We consider method8 to minimize scepticism about the likelihood of provision below when we discuss misspecification of the probability of provision. Likewise, we consider ways to minimize protest zeros given by those who regard government a8 wasteful below under item nonresponse bias. In this atudy, we had no alternative to identifying the Federal government as the provider of improved water quality. People's views about the Federal government were an important factor in promoting protest zero responses despite our efforts, described in the next chapter, to overcome these objections.

Implied Value Cues

Implied value cues occur where the respondent anchors his or her WTP amount on other value8 presented in the CV instrument or implied by it Instead of on the worth of the good itself. The potential for this important type of bias stems from the teohniques wed by CV researchers to reduce the nonresponse rate to the WTP questions. When the WTP amount is elicited by simply asking respondents how much they are willing to pay for the good in an open ended question, many respondents find it difficult to offer precise dollar values for public goods which they are unaccustomed to price, such a8 improved freshwater quality in the nation's lakes, rivers and streams. However, when respondents are offered a context or framework for valuing the good, the number of nonresponses decreases markedly.

The four method8 which have been used by CV practitioners in order to achieve an ● ooeptable response rate are: (1) the bidding game, (2) multiple choice questions or contingent ranking (Desvousges et al., 1983), (3) the payment card, and (4) ● ooeptance or rejection of a single WTP request without bidding where the amount proposed is systematically varied l oroes the sample (Bishop and Heberlein, 1979). Although each has one or *more particular form of* potential bias associated with it, we restrict our comments here to the two most commonly used elicitation methods: the bidding game, which we rejected for uae in this study, and the payment card, which we adopted. Each is subject to a different potential bias.

Starting Point Bias In the bidding game format (Randall et al., 1974), the interviewer proposes a starting bid and asks the respondent if he or she is willing to pay that amount. If the answer is yes, the interviewer increases the amount by some fixed increment and repeats the WTP question; if the answer is no, the interviewer reduces that amount by some fixed increment and repeats the question. This prooeea oontinuee until the yeas change to a nay or vice versa and the amount the person is WTP is narrowed down to a fairly small interval . The potential for bias lays in the likelihood that the Initial starting bid will suggest a value for the good to the respondent. Thus even if a respondent rejects the initial bid, starting points well above the respondent's true WTP will tend to increase the revealed WTP, while starting points well below it will tend to decrease it (Thompson end Roberts, forthcoming). Kahneman and Tversky's (1974) psychological experiments demonstrate that this effeot occurs under fairly general oonditione when the value in queation is not well defined or not frequently considered by the respondent. A preponderance of teats in CV surveys have shown that starting point bias is a very real phenomena when the bidding game format is used and

that this bias is often large relative to the final WTP bide (Rowe et al., 1980; Boyle, Bishop and Walsh, forthcoming).[3]

Range-restriction Bias The bidding game's vulnerability to starting point bias led us to develop the payment card as an alternative elicitation procedure. Although this method, which is discussed at length in Chapter 5, has the potential for other types of bias, the risk of bias from these sources appears to be more manageable. One type of potential bias from payment oarde is relational bias which we will discuss ahortly. The second is the possibility of range-restriction bias.

Although no one starting point is identified on a payment card, the information on the card may restrict the range of the respondents' WTP amounts in the following ways: (1) the maximum amount on the card may be lower than the maximum WTP of some respondents, and (2) the amounts shown on the card may not include the amount the respondent is WTP. The bias from the first restriction is fairly easy to avoid by using a high enough value as the upper anchor of the payment card anchor. The bias for the second is more subtle and difficult to minimize. For example, if the true WTP of a significant number of respondents falls into the gap between the zero and the first positive amount on the payment card respondents, despite the injunction to "choose any amount in between," may feel constrained to choose between zero and an amount greater tban their true WTP thus providing little information about the shape of the

---

3. The typical test for starting point bias has been to perform an analysis of variance or run a regression of the form:

$$WTP = a + bS + e \qquad (1)$$

where WTP is an n x 1 vector of revealed WTP, S is a n x 1 vector of the starting points used (a dummy variable in the ANOVA case), a and b are coefficients to be estimated, and e is a vector of error terms. Carson, Casterline, and Mitchell (1984) argue that more complicated formulations need to be considered as the reaction functions to starting points above and below the respondent's true WTP are likely to be different and non-linear.

underlying demand curve. Alternatively, even if the WTP amount is greater than the first amount above zero, people tend to pick either one of the number on the payment card or a round number such as $10 or $25, instead of valuing the good by amounts such as $17 or $23. Provided the card provides a large enough array of amounts, and the sample is sufficiently large, this type of rounding off should not cause much trouble in computing means or medians. The implications of such rounding off for more 8ophietlcated multivariate analysis are not well known although we do not believe they are serious. We should also point out that the tendency to choose round numbers or payment card interval points is more of a problem when the revealed WTP amounts for the good in question are likely to be concentrated in a fairly small interval, than when the good is a broad national program such as the water quality levels valued In this study.

Yea-saying is the tendency of some respondents to agree with an interviewer's request regardless of their true views (Arndt and Crane, 1975).[4] This form of bias affects both the bidding game[5] and the format where a single asking price is proposed to the respondent for a given level of the good and the respondent is asked whether or not he or she is willing to pay the price.'

---

4. Nay-saying may occur if the respondent feels the desired or normative response is a no. This may frequently occur in studies using the WTA format.

5. In the bidding game, although the respondent does have to respond "no" at some point in the iteration sequence, the yea-saying phenomena suggests that bidding game estimates may be biased upward.

6. Yea-saying bias is potentially a major atumbling blook to the use of the single price format whose simplicity otherwise makes it the method of choice for use in mail or phone interviews. The problem is that there are no ways to identify the nonrespondents such as those who, in regular CV studies, are identified as giving protest zeros or who are defined as outliers and removed from the core data aet . Pare, yea-saying may be seen as akin to the biometrician's problem of how to estimate the effect of a stimulus against a non-zero background, Hanemann (1983) has considered this problem in a CV

(Footnote continued)

It does not propose a potential problem for studies using the payment card method.

Relational bias, the last of the implied value cue biases, occurs when the described relation of the public good being valued to some other good implies a value for the first public good. This bias may occur when values for other goods are used as perceptual aids on payment cards or on visual display handouts such as our water quality ladder or risk ladders where the risk from aeveral different activi ties are shown. In our pretests for this study, we explored, qualitatively, respondents' sensitivity to the position of the three levels of' water quality on our water quality ladder. It did not appear that respondents were so sensitive to the locations that a small miscalculation of the correct position of these levels would -affect our findings. In an effort to help respondents to better understand how they pay for public goods, we " anohored " our payment cards by identifying those amounts on the card which represent how much the average household In each income category was paying for certain other representative public goods. We deliberately did not include any environmental goods as anchors for fear of relational bias. Relational bias is not indicated for our anchors as shown in chapter 5 which discusses an experiment we conducted to test for this bias.

Misspecification of Market Scenario

The nature of CV questions is such that they are very difficult to word so as to eimultaneouely meet the requirements of economic theory end the cognitive needs of nonexpert citizens, such as the proverbial poorly motivated respondent with lees than a high school education, who pauses briefly to contemplate the

(continued)
context. In marketing research, revealed willingneee-to-purchase is typically discounted in aBayesian or ad hoc fashion based upon past observed outcomes, an option nc: readily available In CV studies.

matters described to him or her by the survey Interviewer. There are numerous ways in which the hypothetical market proposed In CV studies can be misspecified in the sense that the market scenario is either incorrect from the standpoint of economic theory ,or is so poorly described that the respondent perceives the scenario to be different from that intended by the survey designer and therefore values what amounts to a different good. The end result of such market scenario misspecifications is biased WTP amounts for the public good in question. Cur efforts to minimize these types of misspecifications in the design and pretesting of our questionnaire are considered at length in chapter 4.

Payment vehicle misspecification occurs when the intended payment vehicle differs from the one perceived by the respondent. We believe our vehicle is reasonably free from problems on this score.

Budget constraint misspecification poses a greater potential problem. CV studies intend that respondents WTP amounts should take Into account the available income of the appropriate unit, typically the respondent's household. One type of misspecification occurs when respondents' think in terms of their personal income rather than their household's income. Another type is when they give an unrealistically high amount because they fail to treat the exercise seriously enough to try to understand how much the good is worth to them if they had to pay the amount. Various techniques have been used to assess and/or minimize this type of bias. In one field experiment, Schulze, cummings and Brookshire (1983, chapter IV) went so far as to have respondents reveal their net monthly household income and allocate it between five categories prior to valuing the preservation of air visibility. After giving their bide, respondents were asked to indicate which expenditure category would be decrease2 in order to finance their contribution. In the water benefits

survey, we sought to maximize our respondents' understanding of the financial implications of their expressed WTP amounts by having the Interviewers add up the amounts each respondent gave for each level of water quality and remind them of the total implications of their payments. We also wed annual taxes instead of (easy) monthly payments.

A third, and opposite type of budget constraint misspecification, occurs when respondents assume that the amount they give is meant to be an addition to their present taxes when, in fact, they are already paying some amount for the good in their present taxes. In this case respondents unwittingly introduce a stricter budget constraint than In fact exists. The present study is, to our knowledge, the first to attempt to deal with this bias. First, we explicitly inform the respondents in the early part of the scenario that they are already paying for water quality in taxes and prices and ask them to imagine that this money is refunded to them and they can determine their future payments. If misunderstandings persist on this point, they should become apparent to respondents at the point in the interview where they are told what their current payments are and they can revise their amounts accordingly.

Amenity misspecification occurs when the perceived good being valued by a respondent differs from the researcher is intended good. The findings of research conducted by cognitive psychologists on how people (both lay people and experts) use information to reach Judgments offers a basis for supposing that respondents may havedifficulty in comprehending certain types of situations in tha way intended by the researcher. For example,according to the representativeness heuristic (Kahneman and Tversky (1972; 1973) people don't treat all the information in a given situation equally, but focus instead on the most representative aspect. Applied to CV acenarioe, this heuristic suggests two types of amenity misspecification -- geographical and benefit --

which can result In part-whole bias. In the geographical case, someone who is asked to value a local environmental amenity would focus on the amenity and imperfectly comprehend the locational aspect of the scenario. The resulting values would be given for the good as a whole instead of being limited to to the geographical part. Similarly, asking respondents to value sub-types of benefits such as option, existence or bequest values may exceed their cognitive capacity to separate these dimensions from the overall value they have for the good.

In this study we attempt to minimize geographical and benefit part-whole bias by first asking the respondents to give, their total value for the country's water quality. Our measure of state level benefits, employs a decomposition strategy where respondents are asked to divide the previously offered total WTP amount between their state and the rest of the country. This strategy minimizes part-whole bias because it focuses the respondent's attention on one aspect of the scenario at a time. We did not attempt to have respondents deoompose their overall WTP amount into benefit subcomponents because we are skeptical of respondents' ability to reliably make the required distinctions between the rather abstract concepts of, say, option and existence benefits. We did estimate a lower bound for the nonuse component of the benefits in chapter 2, but we did this by internal analysis based on their recreational use of freshwater rather than by asking the respondents to make this es timate.

Another and important part-whole issue is whether our respondents are valuing only water quality or whether they are valuing water quality as a surrogate for environmental quality more generally. This latter response, by capturing nonwater benefits, would obviously bias upwards our water quality values. The decomposition technique is inapplicable in this case because of

the difficulty of clearly describing joint quality improvements. Wc attempted to counter this possibility by, among other things, explicitly reminding the respondents that air quality (by far the most costly non-water environmental program) was not included in our study and that they should value water quality under the assumption that air quality would remain roughly the same. In order to test for the presence of this type of part-whole bias, we conducted the experiment described in chapter 1 where we told half the sample what they were already paying for both air and water quality and gave them the chance to revise their water WTP amounts. The results. of the experiment were reassuring because the differences between those who experienced this treatment and those who were only told what they were paying for water quality were not statistically significant.

A rather different type of amenity miaapecification occurs if people do not accept the property right to the good being valued which the researcher attempts to confer on them. We believe property right miaapecification is a key factor behind the vary large number of protest zeros and infinite values elicited by CV studies which ask people how much they are willing to accept (WTA) in payment for giving up some amenity (Hammack and Brown, 1974; Brookshire, Randall and Stoll , 1980). According to economic theory, the WTA and WTP formats should yield equivalent anawera. This ignores the fact that the perceptual reality of these situations is very different for many respondents. The notion of paying for the provision of a public good involves a oollective right to the good which is understandable to moat people because of their familiarity with user fees and tax payments. The property right Implied by the WTA format is another matter because of the widespread belief, baaed on strongly held cultural values, that It is “wrong” to be paid to allow more pollution or to give up one’s right to have a hunting license.

Furthermore, for pure public goods such as air and water quality, people may be unable to Imagine that they personally have a property right to the good and therefore reject the idea of selling something they do not own.[7] Recognition of these problems have lead CV researchers' studying environmental amenities to restrict themselves to the WTP format which we uae In this study.

Probability of provision misspecification occurs when the perceived probability of provision differs from that probability intended to be conveyed by the scenario. Survey designers usually try to convey the idea that the level of the public good being valued will definitely be provided if enough money is raised. If skeptical or cynical respondents discount this certainty, the good will be undervalued. For example,researchers who ask respondents to value reductions in risk levels such as those posed by hazardous waste sites or nuclear power plants,face the difficult task of convincing their respondents that these reductions will in fact occur if the government program being valued is implemented. It was our judgment that this type of misspecification was not a threat to this study because the discernable and well publicized progress that has already been made In improving freshwater quality should give the government credibility In this area.

Context misspecification is the final type of misspecification bias error to be considered. The relevant context for a CV interview includes the physical setting -- location, time of day and year etc. -- and the setting created by the material in the interview which precede the CV scenario. An appropriate context will prepare the respondent to give serious consideration to the questions in a manner conslatent with the intention of the interview. Context misspecification occurs when the setting of the interview either

7. The WTA format also promotes budget constraint misspecification since the amounts are not income bounded in the way that WTP amounts are.

Introduces bias or fails to produce the desired context.

Our scenario waa preceded by a series of attitude questions which was designed to introduce the following context:

- a That water pollution la only one of a number of public goods, some environmental, some not, which the respondent la currently paying for (q. 1-3).
- That envirnmental goods involve a tradeoff between cost and degree of improvement (q. 4).
- A realization. that water pollution is oauaed by a number of different factors (q. 23).

Another sequence of questions probed the respondents houaehold'a uae of freshwater for recreation. Eliciting this information prior to the valuation sequence reminded the respondent of these uaes and whether or not the respondent and other members of his or her household actually used freshwater for these purposes or not. The wording of these questions was designed to present environmental goods and support for environmental protection in a neutral context in order to avoid motivating socially desirable pro-environmental reaponaea.

Aggregation Bias

These biaaea involve error introduced in the course of aggregating the WTP amounts, either across individuals to obtain a population estimate, or across sub-categories of a benefit to obtain a total benefit estimate. Two of these aggregation biaaea have to do with either imperfectly sampling the population (sampling design bias) or with systematically failing to obtain interviews from one or more categories of sampled respondents (response rate bias). A description of our sampling procedures and the weighting scheme we used to minimize sampling bias is presented In appendix B. The other two types of aggregation bias -- item noresponse and sequencee -- pose a significant problem for many CV surveys, including this one.

The first., Item nonreaponae bias, la in many ways similar to nonresponse bias and the correction methods (Kalton, 1983; Panel on Incomplete Data, 1983) are frequently the same. Minimizing item nonresponse and correcting for its presence la particularly important in CV surveys beoauae the crucial WTP questions often have higher item nonresponse rates than are normally found in social surveys. Whereas an item nonreaponae rate of 8 to 10 percent would be regarded as high in an ordinary opinion survey,rates of 20 to 30 percent or even higher are not uncommon in CV surveys where: (1) the sample is random and therefore includes people of all educational and age levels, (2) the scenario is complex, and (3) the object of valuation is an amenity such as air visibility which people are not accustomed to valuing in dollars. Moreover, certain subgroups, such as people with low levels of education, contribute disproportionately to these high nonreaponae rates. Up to a certain point, a higher than usual level of nonresponse to the WTP questions in these circumstances is acceptable and even desirable since heroic efforts to encourage reluctant respondents to answer the WTP questions are likely to result in unreliable or badly biased answers.

Item nonresponaea fall Into four general categories: (1) don't knows, (2) refusals, (3) protest zeros, and (4) responses which fail to meet an edit for minimal conaiatency . In a well designed CV study, the first three categories usually constitute the bulk of the item nonresponses and it is possible through questionnaire design to influence the distribution of nonresponses across these categories. Protest zeros are perhaps the moat troublesane category as it is necessary to distinguish them from true zero bids given by respondents who prefer to forego the good in question rather than to have to pay for it. In this study we experimented with a more thorough than usual probe of why people gave zero amounts which Incorporated oounterargumenta in an attempt to convert

some of the protest zero bidders to genuine bidders.

The responses set to "missisng" during the consistency edit consist primarily of very poor respondents who give WTP amounts which represent an implausibly large percentage of their income, and some upper income respondants who indicated zero or low WTP despite the fact that they gave answers to other questions In the survey indicating a strong demand for the good. various techniques can be used to perform this edit ranging from deleting observations on the basis of the ratio of WTP to income (Tolley and Randall 1983) to wing the regression outlier approach adopted by Deavouagea, Smith and McGivney (1983). As the criteria for defining an observation as an outlier are judgmental, it is Important to explicitly describe for each case the reasons why the WTP amount is rejected as invalid. Appendix E presents this Information for the 26 outliers deleted In this study.

The bias introduced by the first three types of aggregation bias is often aaaeaaed by comparing the relevant distributions of the data with the actual cenaua distributions for key demographic characteristics. As noted in chapter 1, we interviewed (self-identified) houaehold heads because they are better able to speak authoritatively. about the household's values for water quality than other household members such as dependent children or senior citizens who are living with their children. Table 1 presents data on six demographic categories for the 1983 U.S. houaeholder population and several versions of our sample. Column A shows the characteristics for the 811 individuals in the complete data act,B for the core sample who gave usable answers to the WTP questions and C for the sample after adjustments were made for the several types of nonresponse. These data allow ua to assess the bias due to (1) sample design/nonresponse and (2) item nonresponse in this study.

Looking first at A, we see that our sample matches the census distribution

Table 1 DISTRIBUTION OF TOTAL AND CORE RESPONDENTS AND NATIONAL POPULATION ON KEY DEMOGRAPHIC VARIABLES[a]

Sex[1]	Census	A	B	B-A	C
Male	48%	44%	46%	+2	44%
Female	52	56	54	-2	56
Race					
Black	11		0	-1	
Other	89	9:	92	41	9:
Education					
Less than H.S. Graduate	28	25	19	-6	31
H.S. Graduate	36	36	37	41	34
Some College	16	22	244	42	20
College Graduate	20	17	21	44	16
Age					
	7				
18 - 24	23	11	12	41	11
25 - 34		23	26	43	23
35 - 44		18	18	0	16
45 - 54	19	14	13	-1	13
	15	16	14	-2	16
55 - 64	16	19	16	-3	2 1

Before Tax Household Income[2]	Census (1960)	a	B	B-A	c
O - 14,999	42	40	36	-2	43
15 -21,999	26	26	27	41	26
25 -19,999	27	27	28	41	26
50,000 +	5	7	8	+1	6
Region					
Northeast	21	21	20	-1	21
North Central	25	25	27	-2	26
South	34	34	31	-3	34
West	20	19	22	+3	18

A = Total sample before weighting, N = 813
B = Core sample - those with usable WTP amounts before weighting, N = 564
C = Adjusted sample after **weighting** and imputing values to those who did not answer the **WTP questions, a = 1,019**

[a] Unless otherwise indicated, national data are 1983 U.S. Census estimates for "householders" from U.S. Census Bureau, Household and Family Characteristics: March 1983, Current Population Reports, Population Characteristics, Series P-20, No. 388 (U. S. Government Printing Office, Washington, D.C., 1983).

1. This distribution is calculated from the 1984 report which lists 50 million married couple families, 11.5 l llllm "other family" or "nonfamily" households with male householders and 22.5 with female householders.

2. Before tax household income from U.S. Bureau of the Census, Statistical Abstract of the united States 1984 (U.S.Government Printing Office, Washington, D.C., 1984).

closely on sex, income, and region. There is a small degree of bias on age, education and race. Our sample of household heads has four percent more householders age 18-24 than the general population and slightly fewer people over age 65. It also has two percent fewer blacks and the fit on education is off slightly although not In any one direction.

A comparison of column B, the core of usable responses, with the total sample in A, shows the demographic characteristics of those among the respondents who did not provide usable unusable WTP answers. The difference scores are consistent with the hypothesis that. people who are leas able to cope with the cognitive demands imposed by the valuation process tend to be nonrespondents to the WTP questions, as there are strong education and age effects. The difference scores, B-A, show that those with leas than a high school education and those age 65 and over were least likely to be able to answer the WTP questions.

Column C shows how ORC's weighting procedure affected the distribution. ORC's procedure imputed the minimum deviation from five the census categories of race, region, education of head of household, household income, and number of people In the household. The effect of these prooedurea on the pre-weighted sample's deviations from the variables shown in table 1 is minimal, owing to the tightness of the original fit. Poll data on environmental preferences suggests that any bias lntroduoed by the age category and race discrepencies between column C and the oenaua data, will be cnacel each other out. Although support for environmental protection la remarkably broad baaed across demographic categories (Ladd, 1983) blacks are somewhat 'leas supportive than other categories (Ladd, 1983; Mitchell 1980) and young people are somewhat more supportive.

Given the level of Item nonreaponae in CV surveys and the fact that the

poor, leas educated,elderly and those with no interest In the public good being valued are disproportionally represented among the nonrespondents, there is a great need for methods which can adequately Impute the missing WTP amounts.[8] CART, (Breiman et al., 1984), the procedure wed In this study, appears to work beat. After each case was weighted, everyone in the 30 percent of our respondents who did not give a usable WTP amount had one Imputed to it by the CART procedure baaed on the WTP amounts given by those in the sample who moat resembled the respondent's demographic and attitudinal profile.

The second type of aggregation bias to be discussed here is sequence bias. Hoehn and Randall (1982) have identified a sequencing effect which occurs when the value of a particular good or policy depends on the order in which it is valued in relation to the other goods or policies in the sequence. Asked to value environmental Improvement A, and then environmental Improvement B, people will offer higher values for A than when the sequence is reversed. These effects were empirically demonstrated in studies of regionally specific air pollution benefits where Randall,Boehn, and Tolley (1981) found respondents were willing to pay more than five times more for a specified atmospheric visibility program for the Grand Canyon area when this was the only visibility program they were asked to value compared to the amount they were willing to

---

8. This la one of the moat active areas of current research Into nonsampling errors (Panel on Incomplete Data, 1983). Carson (1984) compares several methods of imputing missing WTP responses for a CV survey of rational freshwater benefits including the hot deck, the EM procedure (a maximum likelihood approach) ,CART (a nonparametric tree-structured approach), and no imputation. All of the methods suggest that failing to impute values for the missing data in that particular CV survey results in approximately a 25% overestimate of mean WTP. The hot deck technique (Bailar and Bailar , 1978) appears to work reaaonably well, the only problem being how to choose the Imputation classes and the fact that missing WTP responses tend to run together in particular sampling points having the demographic characteristics mentioned above. The EM Procedure (Orchard and Woodbury, 1972) also appears to work well except that it produces some negative WTP estimates which would have failed an edit.

pay for this same program when it was valued last in a three part sequence.[9]

Because goods are substitutes and complements, the sequencing effect itself is an understandable economic phenomenon end the different. values are not biased. Thua, respondents will value cleaning up the first lake in their area more than the second lake and so on. However, and here is where aggregation biaa occurs, If separate CV studies value each lake, each is necessarily treated as the first lake, and adding the values to get the benefits of cleaning up all of the lakes in the area will result in over valuation. Thus, sequence bias occurs if separately measured values for several component policies (or subcomponent goods) of a program are combined to produce a total value for the program (or good).' Sequence bias la not unique to CV studies; it poses a serious problem for other benefit measurement techniques, such as travel coat and hedonic pricing, as well. Because we value the entire water pollution control program for the United States in the water benefits study, no aggregation was necessary to arrive at a value for the national program and sequence bias was minimized.

Factors Affecting the Variance of Willingness to Pay Amounts

In the preceeding section we explored the possibility that our findings might be invalid because of error induced by one or more sources during the interviews. In this section we consider the Issue of reliability. It is sometimes argued that respondents are not sufficiently motivated to expend the effort necessary to give meaningful answers In CV studies (Freeman, 1982; Feenberg and Mills, 1980). According to a critic of en early CV study, "ask a hypothetical question, get a hypothetical answer" (Scott, 1965) where the term hypothetical is used in the sense of nonsensical or useless. We have already

9. The other two parts of the sequence were air visibility programs for Chicago and the United States east of the Mississippi.

pointed out that one consequence of a situation where respondents are asked to value something which is not meaningful to them is for them to give biased answers. Alternatively, if respondents do not rely on extraneous aspects of the scenario for response clues, another outcome is WTP amounts which are notlcably random, characterized by high variances and no statistical relationship to those respondent characteristics which we would normally expect to be predictors of their value for the good In question.

In order to explain the difference between the WTP amounts given by respondents, it is necessary to distinguish systematic variance from random variance. We posit that a respondent's WTP for a public good la described by,

$$WTP = f(X,B) + e(Z) \qquad (2)$$

where X is a matrix of the respondents attributes' and B la a matrix of unknown parameters. The e term is a random variable which is influenced by a number of variables Z (X can be a subset of Z) including the respondents' familiarity with the public good, their ability to conceptualize purchasing different quantities of the good, and the design of the market scenario presented by the survey instrument. The problem is how to decrease the noise level e (Z) relative to the signal f (X,B) and how to extract the signal.

Extracting the signal la not an easy problem as the function, f(X,B) is likely to be highly non-linear with poor people having little extra money to pay for public goods In spite of their needs, uses, or attitudes toward that public good, while the wealthy may exhibit fairly complex willingness-to-pay patterns baaed upon their attributes. These obstacles notwithstanding, the WTP amounts given by respondents in a number of CV studies (see Schulze et al., 1981; Mitchell and Carson, 1984), including the present study, are well explained in regression analysis, are reasonably compatible with amounts implied by other economic valuation techniques (where those available), and/or

have been replicated successfully on independently drawn samples. Since the techniques for estimating the parameters, B, are no different from those commonly used in moat social science disciplines, we concentrate hire on how e(Z) can be influenced by the survey design. The trick la to do this without Introducing Incentives for respondents to bias their answers.

There are several ways to reduce e(Z) without much risk of biasing f(X,B). Sane of these have to do with anticipating the mistakes that people are likely to make in answering the questionnaire and providing opportunities for respondents to learn about and correct this misunderstanding. For instance, there are two ways to ask about WTP for different levels of public good. One la to ask for the marginal WTP to go from each specific level of the good to a higher level. The other is to ask for total WTP from the base level to each level of Interest. In our pretests for this study, we discovered that no matter which approach was used, a significant number of the respondents answered the WTP questions as if they were asked in the opposite way. We found that higher quality WTP data were obtained by designing the questionnaire so the respondents were shown their answers immediately after the sequence of WTP questions (summing if the marginal method was used) and giving them the opportunity to change or reallocate their WTP amounts at that point.

Procedures to separate true zeros from protest zeros and to recover protest zeros are also important. The reasons for giving a zero WTP amount can be probed using procedures such as those used in this study, which offered those giving protest *zeros* ● ddstlonal information in an attempt to overcome the objections which lead to the protest zero. Where such respondents can be induced to give WTP amounts, they should be better estimates than if their values were counted as missing and substitute values were imputed. We were able to identify the protest zeros in the present study, but our efforts to

recover them as uauable reaponaea, which we describe in the next chapter, were not particularly successful.

Two other helpful ways to minimize e(Z) must be implemented with care in order to avoid biasing f(X,B). The first involves using the early part of the questionnaire to engage the respondent in thinking about the good, on the assumption that the more familiar a respondent is with the good being valued, the leas random his or her valuation is likely to be. This can be done by the use of questions which explore the respondent 'a knowledge, use, and attitudes toward the good, and by posing tradeoffs between that good and other public and private goods. We used both techniques in this study. The potential bias posed by this procedure would be a form of context bias, where the questions highlighting the good would lead respondents to overvalue or undervalue it. As noted earlier, we attempted to avoid such a bias by balancing the introductory questions so that they introduced material which supports both pro and anti environmental positions and by wording the questions in as neutral a fashion as possible. We also paid careful attention to whether our pretests showed evidence of context bias from these questions.

A second way to improve the quality of the WTP responses in CV studies of relatively unfamiliar goods,is to provide respondents with relevant examples of the value of other goods. The assumption in this case la that because most respondents have little knowledge about the amounts they are currently paying for public goods,this Information would provide a helpful oontext. Certain types of anchors have the potential to induce bias, however. Chapter 5 discusses the pros and oona of different types of payment card anchors in some detail.

Reliability and Validity of the Water Benefit Estimates

Having described the procedures we used to minimize the bias and random

variance in our estimates, we now direct our attention to quantitative evidence for their reliability and validity. Both dimensions can be assessed in a variety of ways, a number of which are applicable to this study.[10]

Reliability

Although we do not have"test-retest" data for our sample,[11] there are several indications of our data's reliability. First, as noted in chapter 1, the fact that our WTP amounts are consistent with those measured in our 1980 pilot study suggests a certain stability in valuation which la consistent with reliability. Second, the satisfactory amount of variance explained in our estimations is strong evidence that respondents were not randomly answering the WTP questions. Third, the number of those who changed their responses, when given the opportunity to revise them after the first valuation round, was high enough to suggest that our respondents were not afraid to admit that their first answers were inaccurate and yet was not so high as to indicate widespread confusion about the valuing process. Likewise, when we told the respondents what similar households were paying for water quality, the revisions were relatively modest and made sense. Finally, we undertook a systematic search for respondents whose responses Indicated that they misunderstood the valuing process or that they gave amounts which were not income bounded in a meaningful way. The number of respondents whose answers we considered to be outliers,

---

10. Much of the work on measurement theory has been done by psychometriclana (see Lord and Novick, 1968 for an overview). the following works discuss validity and reliability from other disciplianry points of view: Lansing and Morgan (1971) for economica, Carmines and Zeller (1979) for political science and sociology and Borhrnatedt (1983) for sociology and survey research.

11. This would require the reinterviewing of all or some of the sample to determine the stability of their WTP amounts at two points In time.

12. For a discussion of this process and a descriptive list of each outlier see appendix E.

and therefore removed from the core of usable WTP reaponaea, was small enough (26 or 3 percent of the sample) to suggest that moat of our respondents who answered the WTP questions were able to cope with the instrument, despite its complexity.

Validity

As noted earlier, the concept of validity, or whether an indicator actually measures the concept it is intended to measure, is moat clearly and unambiguously demonstrated if the indicator can be used to predict a suitable criterion. A plausible criterion for our water quality estimates would be the results of a national referenda where voters would be asked whether or not they approvated a measure to meet the fishable goal at an indicated coat to each household in taxes and higher prices. If valid, our data should successfully predict such a referenda provided It was conducted within a reasonable length of time after our survey. Unfortunately, although it is possible to conceive of possible criteria such as this referendum, they do not now exist for the good measured in this study.

It might be thought that we presented evidence for predictive validity in chapter 1, where we showed the close fit between Vaughan and Russell's travel coat baaed value and our value for the partial Improvement of the nation's minimm level of freshwater quality from 95 percent fishable to approximately 99 percent fishable. The Vaughan-Russell's value's appeal as a criterion comes from the fact that it is behavior-baaed and many economists place more credence on this type of data than on survey responses. Unfortunately, behavior-baaed measures are themselves prone to to numberous forma of error (Smith, 1984) and cannot qualify as true criteria. Thus, while correlations between behavior-based and CV-baaed measures are evidence for "convergence validity, the absence of a strong correlation In not sufficient evidence to argue that

one of the two measures la leas valid than the other.

In the absence of definitive teats for predictive or criterion-baaed validity, CV studies need to assess two other types of validity; content and construct. The content or face validity of a research instrument depends on the extent to which it reflects a specific domain of content (Carmines and Zeller, 1979). This type of validity has been of particular concern to psychometricians who use It when validating attitude scales. It la also Important to CV studies because it is a necessary condition for a CV study's validity that it's scenario be:a) conaiatent with the requirements imposed on the study' by economic theory and b) likely to be understood by the respondents In the way intended by the researcher.

Two types of evidence are relevant to aaaeaa a CV Instrument's content validity: the findings of experiments designed to test whether the correct domain was understood by the respondents, and qualitative examination of the research instrument to see if it appears to present the correct information end ask the right questions In an understandable fashion. The experiment we conducted to teat whether we captured only water quality values (as intended) or environmental values more generally, was designed to provide evidence about what appeared to be a particularly threatening misspecification problem. Although far from definitive,our finding In support of the hypothesis that people were not valuing both air and water quality la important evidence for this study's content validity. As it is not possible to conduct experiments to teat for every possible type of misspecification, the face validity of our findings must also be qualitatively assessed by consulting appendix A and examining the degree to which our questionnaire's scenario meets the two criteria mentioned above. To the extent that the questionnaire incorrectly specif ies the content domain in either respect, the validity of our findings is

suspect. In an effort to ensure this type of validity, we had the instrument reviewed by several outside economists to see 'if the scenario was properly specified. We also conducted the extensive pretesting program described in the next ohapter to make the instrument as understandable as possible. The degree to which we succeeded In this endeavor is a matter of judgment and the interested reader will have to decide this matter for him or herself.

Presuming that face validity is present, It is necessary to assess a measure's construct validity. This type of validity is concerned with the extent to which a particular measure relates to other measures in a manner consistent with theoretically derived hypotheses (Carmines and Zeller, 1979). If our instrument is valid, we expect respondents to give rational answers to the WTP questions so that, for example, their WTP amounts will be consistent with their ability to pay and with their environmental preferences. In chapter 1, we described a theoretical model which we were able to estimate in part by regressing measures of income, recreational use of freshwater, and attitudes towards spending public money for water pollution control on $WTPTOT_R$. Several aspects of that estimation support the construct validity of our findings. The first is the high (for a cross-sectional study) level of explained variance obtained by this parsimonious model. The second is the model's specificity. When the predictive ability of the attitude item measuring support for spending on water pollution is compared with a similar item for air pollution, only the water pollution item enters the equation significantly.[13]

The available evidence[14] supports the validity of our WTP measures;

---

13. Successfully estimating WTP amounts thus provides evldenoe that they are reliable and valid in **this sense.**

14. Which includes the qualitative findings of our pretests presented in the next chapter.

respondents do seem to be valuing water quality benefits when they give us their WTP amounts. This said, we must point out that our WTP amounts may be valid by these standards without necessarily being very accurate. If everybody's WTP amount was 50 percent below or above their true amount, our refressions would predict just as well and our questionnaire would have the same degree of face validity as if our estimates were highly accurate. The confidence intervals we present in Chapter 2 do not speak to this issue because they only reflect sampling characteristics.

Our approach to this issue has been to design the study so our estimates represent a credible lower bound to water benefits. Accordingly, to the extent that our estimate is inaccurate, it is very likely to under rather than overestimate the true benefits. This is so for a number of reasons. First, as noted at the end of the preceeding chapter, our design decisions were conservative in the sense that when faced with alternative procedures with possible biasing consequences, we chose the alternative whose effect would be to minimize the incentives to overreport willingness to pay. Second, the sampling bias that remains uncorrected for is likely to lower our estimates below their true level. Third, that the aggregate WTP amounts increased when we provided Information about what the respondents' were actually 'paying for water quality, and Increased still more when we provided a further Incentive for raising them, confirms our assumption that respondents are risk averse when revealing large WTP amounts. For national water quality, a significant number of respondents are willing to accept higher payments when they realized what they said they were willing to pay was lower than the current amounts they are paying and relatively few want to reduce their amounts in the reverse situation.

Conclusion

Conclusion

This chapter presented a framework by which the validity and reliability of CV studies can be assessed. Despite the large number of potential nonsampling errors Identified in this chapter, the contingent valuation method remains an important and viable method to measure the benefits of many nonmarketed goods. CV is virtually the only method capable of measuring most nonuse benefits. While other methods are capable of measuring use benefits, they are not necessarily superior for that purpose to a well designed and executed CV survey.

Our purpose in developing thisframework was to Identify the potential biases we needed to address In this study. In this and the preceeding chapter we describe the approaches we adopted to minimize bias from these sources and assess the evidence for the reliability and validity of our estimates. These data suggest that the adjusted WTP_R estimates are a credible lower bound for national water quality benefits.

Part II METHODOLOGICAL ISSUES

4. DEVELOPMENT OF THE WATER QUALITY INSTRUMENT

In this chapter we review the evolution of our research instrument which took place over a period of three years and involved several different types of formal pretests. The purposes of this review are twofold. First we wish to provide the reader with the rationale behind about many of our design choices in the hopes that it till aid his or her interpretation of our findings. A CV study is no better than its questionnaire. Second, we hope that our experience in developing this instrument may be useful to would-be CV practitioners who face similar design tasks.

Design Problem

In designing our questionnaire we had to meet the twin criteria we have mentioned before in this report:the instrument had to gather data which met the requirements of economic theory and it had to do this In such a way that respondents could understand the questions and give meaningful responses. Contingent valuation studies differ in the length of their survey instruments. Relatively short instruments can be used when the scenario and the good are well understood as is sometimes the case with local recreational amenities. This was not the oase with our study. The nature of the good we were valuing required us to communicate a relatively large amount of unfamiliar and sometimes complex information in our scenario. For this reason, the second of these two criteria posed the most difficult challenge.

Some examples: For the first time in a CV study, the principal emphasis was on a national rather than a loaal good. This meant that the good was more abstract for many respondents than most local goods. It also raised the potential for respondents' to answer In terms of their local water quality

rather than the nation's overall water quality. Second, unlike the numerous CV studies of visibility or landscape benefits, freshwater quality did not easily lend Itself to pictorial representation. A third aspect of water quality which required explanation was the concept of "minimum" quality. In valuing the fishable and swimmable levels, we were valuing the improvement of only those water bodies which would be raised to these quality levels. Finally, we needed to explain how water quality Iis maintained and improved, both in terms of treatment and payment. Many consumers do not understand the connection between government water quality regulations and price Increases in certain consumer goods.

Design Stages

Our research instrument went through numerous intermediate versions before we arrived at an acceptable final version. The development process can best be summarized as comprising three stages and five principal versions of the questionnaire.

Stage

The first stage included a period of initial development, beginning in 1979 and culminating in the administration of what we will oall Version I to a national sample in 1980. We decided to define the good In terms of the minimum national levels of boatable, fishable and swimmable freshwater quality because these levels are both policy-relevant, since they are used in the Clean Water Act to describe its goals, and because they appeared to be potentially understandable by respondents. After informal pretesting confirmed that respondents did regard these quality levels as meaningful, we developed the water quality ladder for use as a visual aid.

Our intention in developing the ladder was to, convey to people the notion of a range ofdrinking water quality from extremely pure to very impure, to

show the monotonic relationship between the three quality levels we wished them to value, and to distinguish these levels from the level of water quality where It can be drunk directly from a natural water body without harm <drinkable water). It was necessary to relate the boatable-fishable-swimmable levels to water quality parameters with sufficient accuracy so that bias would not be introduced by an incorrect placement of these levels on the ladder. Our colleague, William J. Vaughan, devised a water quality index which we used for this purpose and which is described in appendix C. For reasons stated there, such an index is necessarily crude as It is very difficult to determine the link between scientifically measured quality characteristics and perceived water quality characteristics. We conducted informal pretests to asoertain how sensitive peoples' WTP amounts were to the location of the three water quality levels on the ladder, and therefore how accurate the placement of the levels on the ladder needed to be. Since they indicated that people's WTP amounts were insensitive to small (e.g. 1 rung or so) changes in the levels' relative and absolute location on the ladder, we determined that our ladder was unlikely to be a significant source of bias.

Another major aspect of our design effort at this stage, was our decision to develop an alternative to the bidding game technique which we felt was too vulnerable to starting point bias. Our solution, the anchored payment card, was the subject of an experiment in the 1980 survey. The next chapter describes the development of the payment card format in some detail.

For our payment vehicle, we chose "higher prices and taxes." This vehicle seemed advisable for several reasons. First, it is plausible since this is the way consumers pay for water quality. Second, we felt It would mitigate against strategic behavior because respondents would be unlikely to conceive that their responses to the survey would have an immediate or direct effect on their tax

or price burden. Third, surveys taken in the late 1970s and the results of the various "proposition 13" type referenda, showed strong consumer resentment of taxes. In view of the negative symbolism embodied in the concept "higher prices and taxes," it appeared to us that this payment vehicle would elicit thoughtful responses and that If there was any bias associated with the vehicle, it would be towards under rather than overvaluation of the amenity.

The complete text of Version I's water benefit scenario is presented in figure 1.1 Although- much shorter thsn the comparable portion (the water benefits scenario proper) of our final instrument, there is a great deal of continuity between the two in terms of their basic design elements. This is especially the case with the ladder card and payment cards. A pilot study was conducted in the winter of 1980 when Version I was administered by personal interview to a national probability sample of 1576 people. We chose to conduct a national pilot study, because we were presented with the opportunity to "piggyback"non an existing national survey at a low cost. This arrangement allowd us to test the survey under the conditions in which it would ultimately be administered and to have a large enough sample to conduct a valid test of the payment cards. The circumstances of the survey did not permit us to train the interviewers beyond a minimal set of instructions which they received prior to undertaking the field work. Since our funds did not permit us to have The Roper Organization, who conducted the field work, contribute Its expertise to

---

1. This is the A form of the A - D forms of the scenario. Each form was administered to an equivalent subsample. A, B and C had virtually identical question wording, the only differences were in the format of the payment cards. Version D used a very different format for questions 82ff. Respondents were first told what people of their income group were paying and were asked if they were willing or not to pay this amount each year to raise water quality to level C. Then they were asked how much more they were willing to pay to raise it tc the swimmable level and what amount of money they would be willing to pay to keep It at level D.

Figure 1 VERSION I OF THE WATER BENEFIT CONTINGENT VALUATION INSTRUMENT*

NOTE: INSERT THIS FORM AFTER PAGE 14 OF WHITE "X" QUESTIONNAIRES ONLY AND ASK FOLLOWING Q.79.

80. This last group of questions is about the quality of water in the nation's lakes and streams. Congress passed strict water pollution control laws in 1972 and 1977. As a result many communities have to build and run new modern sewage treatment plants and many industries have to install water pollution control equipment.

Here is a picture of a ladder that shows various levels of the quality of water. (HAND RESPONDENT WATER QUALITY LADDER CARD) Please keep in mind that we are not talking about the drinking water in your home. Nor are we talking about the ocean. We are talking only about freshwater lakes, rivers and streams that people look at and in which they go boating, fishing and swimming.

The top of the ladder stands for the best possible quality of water, that is, the purest spring water. The bottom stands for the worst possible quality of water. Unlike the other ladders we have used in this survey, on this ladder we have marked different levels of the quality of water. For example.... (POINT TO EACH LEVEL: E,D,C AND SO ON, AS YOU READ STATEMENTS BELOW)

Level E (POINTING) is so polluted that it has oil, raw sewage and other things in it, has no plant or animal life and smells bad

Water at level D is okay for boating but not for fishing or swimming

Level C shows where rivers, lakes and streams are clean enough so that game fish like bass can live in them

Level B shows where the water is clean enough so that people can swim in it safely

And at level A, the quality of the water is so good that it would be possible to drink it directly from a lake or stream if you wanted to

Now let's think about all of the nation's rivers, lakes and streams. Some of them are quite clean and others are more or less polluted. Looking at this ladder, would you say that all but a tiny fraction of the nation's rivers, lakes and streams are at least at level D in the quality of their water today or not?

All but a fraction at level D... 1

Not at level D..... 2

Not sure..... 3

81. As you know it takes money to clean up our nation's lakes and rivers. Taking that into account, and thinking of overall water quality where all but a tiny fraction of the nation's lakes and rivers are at a particular level, which level of overall water quality do you think the nation should plan to reach within the next five years or so--level E, D, C, B or A?

A..... 1

B..... 2

C..... 3

D..... 4

E..... 5

Depends (vol.)..... 6

Other (vol.)..... 7

Not sure..... 8

INTERVIEWER: CHECK INCOME IN Q.79 ON PAGE 14 OF MAIN QUESTIONNAIRE. THEN LOOK BELOW TO SEE WHICH SCALE CARD RESPONDENT USES IN QUESTIONS 82 - 84.

IF LESS THAN $9,999 USE CARD A-I

IF $10,000 TO $14,999 USE CARD A-II

IF $15,000 TO $24,999 USE CARD A-III

IF $25,000 AND ABOVE OR NOT SURE/REFUSED USE CARD A-IV

* Version I consisted of four subversions, A,B,C, and D. Versions B and C differed from A only in the configuration of their payment cards.

NOTE: INSERT THIS FORM AFTER PAGE 14 OF WHITE "X" QUESTIONNAIRES ONLY AND ASK FOLLOWING Q.79.

80. This last group of questions is about the quality of water in the nation's lakes and streams. Congress passed strict water pollution control laws in 1972 and 1977. As a result many communities have to build and run new modern sewage treatment plants and many industries have to install water pollution control equipment.

Here is a picture of a ladder that shows various levels of the quality of water. (HAND RESPONDENT WATER QUALITY LADDER CARD) Please keep in mind that we are not talking about the drinking water in your home. Nor are we talking about the ocean. We are talking only about freshwater lakes, rivers and streams that people look at and in which they go boating, fishing and swimming.

The top of the ladder stands for the best possible quality of water, that is, the purest spring water. The bottom stands for the worst possible quality of water. Unlike the other ladders we have used in this survey, on this ladder we have marked different levels of the quality of water. For example.... (POINT TO EACH LEVEL: E,D,C AND SO ON, AS YOU READ STATEMENTS BELOW)

Level E (POINTING) is so polluted that it has oil, raw sewage and other things in it, has no plant or animal life and smells bad

Water at level D is okay for boating but not for fishing or swimming

Level C shows where rivers, lakes and streams are clean enough so that game fish like bass can live in them

Level B shows where the water is clean enough so that people can swim in it safely

And at level A, the quality of the water is so good that it would be possible to drink it directly from a lake or stream if you wanted to

Now let's think about all of the nation's rivers, lakes and streams. Some of them are quite clean and others are more or less polluted. Looking at this ladder, would you say that all but a tiny fraction of the nation's rivers, lakes and streams are at least at level D in the quality of their water today or not? ___

All but a fraction at level D... 1

Not at level D..... 2

Not sure..... 3

81. As you know it takes money to clean up our nation's lakes and rivers. Taking that into account, and thinking of overall water quality where all but a tiny fraction of the nation's lakes and rivers are at a particular level, which level of overall water quality do you think the nation should plan to reach within the next five years or so--level E, D, C, B or A?

A..... 1

B..... 2

C..... 3

D..... 4

E..... 5

Depends (vol.)..... 6

other (vol.)..... 7

not sure..... 6

INTERVIEWER: CHECK INCOME IN Q.79 ON PAGE 14 OF MAIN QUESTIONNAIRE. THEN LOOK BELOW TO SEE WHICH SCALE CARD RESPONDENT USES IN QUESTIONS 82 - 84.

IF LESS THAN $9,999 USE CARD A-I

IF $10,000 TO $14,999 USE CARD A-IX

IF $15,000 TO $24,999 USE CARD A-III

IF $25,000 AND ABOVE OR NOT SURE/REFUSED USE CARD A-IV

(OVER)

further developing the Instrument for national administration, it was administered "as is,""after only light editing.

The results, reported in Mitchell and Carson (1981), were extremely useful. They showed,first, that a national CV study of water benefits was feasible. Cur WTP amounts made sense, the estimations were encouraging, both the water quality ladder and the payment cards proved to work well in the field, and the payment card experiment showed the anchors did not bias the results. Second, the results showed us where further design work was necessary before valid and reliable national water quality estimates could be obtained. The primary problem was that only half of the respondents gave usable WTP amounts. Among other problems was a certain amount of confusion by some respondents about whether we wanted them to give us the total amount they were willing to pay for the fishable and swimmable levels or whether they were to give the incremental amount above the amount they had already given for the lower quality levels.,

Stage 2

In the winter and spring of 1983, further work on the instrument produced Version II. Since the next administration of the Instrument would be in a self-contained survey, it was necessary to add background questions and predictive attitude questions. We also added a more elaborate set of questions to measure recreational use of freshwater by the respondent and the respondent 's household members. (In Version I only information about the respondent was measured.) Other changes and additions included the following:

- A more elaborate description of the concept of minimum water quality.
- A map to help respondents understand that we wanted their values for national water quality.
- In Version I the Initial WTP question for boatable water ended by tellingrespondents: "If it is not worth anything to you, please do not hesitate to say so." Our experience In the pilot study made it clear

When survey researchers speak of instruments that do not "work," they refer to a quality that cannot easily be quantified, and is therefore somewhat elusive. It is also a quality clearly apparent to an experienced interviewer when he or she uses a questionnaire in the field. The concept refers to at least three aspects of a questionnaire. First, an instrument does not work If the respondents find one or more of the questions meaningless or confusing, or if the questions are 90 difficult to answer that too many people refuse to answer or, even worse, give an answer without seriously considering the question. Second, even if the first condition is met, an instrument does not work if the questions do not fit together in such a way that they lead logically from one to the other and from section to section. Third, an Instrument does not work if portion9 of It Involve narrative or explanation that is so lengthy that respondents become bored and restless. In one way or another, version II was guilty on all three counts.

After considerable effort and mutual consultation, Version III was completed and pretested on a nonrandom sample of 100 people who represented the full range of the potential respondents In terms of age, education, income, race etc. Many wording changes were made in this version, the token allocation scheme was dropped because it was too complicated, the order of the items was changed somewhat to Improve the flow, some of the description in the scenario was removed, and the scenario's narrative (e.g. the description of how people pay for watter quality, why they might value it etc.) was broken up by more opportunities for respondent participation. A reminder card for the interviewers was added 90 they could keep straight the dollar amounts the respondent gave for each level and calculate the total. The three interviewers who conducted the pretest were especially chosen for their experience and ability to report;on their experience with the questionnaire.

Two experiments were conducted in the pretest by use of split samples. In the payment card experiment,half the respondents received an anchored and half received an unanohored payment card. A second experiment compared two techniques for having respondents allocate their WTP amounts between in and out-of-state water. One technique had respondents divide a circle into two segments to represent the proportion of the whole they wished to spend for local and nonlocal water quality. The other simply asked them to divide the amount they gave for swimmable water between their state and the rest of the country in dollars or percents.

At the conclusion of the pretest,the Interviewers were debriefed at a meeting which we attended. On the basis of their experience, and our analysis of the data, we concluded that the instrument could be made to work but it needed further improvements. These included dropping the triple ladder card, which was more confusing than helpful, and replacing It with a version of the interviewer reminder card on which respondents would enter the amounts they said they were willing to pay and total them. We felt this approach would help the respondents understand that they were being asked to value the three levels of water quality, that their amounts are incremental ("how much more would you be willing to pay for fishable etc."), and that the total amount for swimmable level is the sum of these amounts. Other indicated Improvements involved reordering some of the questions, tightening the wording of some questions, and breaking up the scenario's description of water quality with still more opportunities for the respondents to participate. According to the interviewers unanimous judgment, the anchored payment card and the non-pie chart division of local and national benefits worked much better than their alternatives.[3] This was an Important finding.

In July and August, 1983, RTI conducted two rounds of interviewing five

respondents each to assess the effect of these changes. The result of all these efforts, Version IV, was and sent to several economists familiar with the contingent valuation method for their comments, which were favorable.

Stage 3

As a result of the work done in stage 2, we hoped Version IV was close to a final version. A further round of revision was yet to come. Several leading survey research organizations, each of which conducts frequent national surveys, were given Version IV and asked to bid on preparing a final version of the questionnaire and administering It in the field. Since a national survey involves numerous (100 or more) sampling points scattered across the country, It is prohibitively expensive to train the interviewers in person. The interviewers are, of course, experienced professionals and are capable of administering very complex questionnaires. Moreover, they receive careful instructional materials, including tapes which go through the questionnaire step by step, and conduct practice interviews, monitored by the firm's' headquarters before venturing forth with a new survey. But even 90, unless a questionnaire is developed to the point where It works smoothly and without oonfusion for most respondents, the potential for confusion and error is great.[4] These firms were unanimously of the Judgment that more revisions and pretesting would be necessary before their interviewers could successfully administer our instrument.

---

3. A reexamination of the notes we made at the debriefing showed that one or more of the interviewers identified problems which did not seem to be of sufficient importance to require major changes but which, when they emerged again in a later round of pretesting, led to important changes in the instrument.

4. A national sampling plan involve9 100 or more sampling points. The expense of bringing the interviewers, who live In these areas, together for personal instruction could not be covered by our budget. The Instructional **program** used for this survey is described in appendix B.

He therefore developed a further pretesting effort with the Opinion Research Corporation (ORC), the firm chosen to do the work. Its key component was a series of interviews which were conducted over a two day period in an ORC facility which permitted each interview to be observed and taped for followup analysis. The interviewees, who were chosen to represent a range of respondent characteristics, were paid for their time. Mitchell participated in this process, which involved reworking the instrument after each interview to address the problems revealed during the trial, and then administering the revised version to the next interviewee. This process continued until all the problem9 were taken care of. Despite Its apparent artificiality, the observed interview prooedure was a very efficient way to determine just where the continuity of the interview was unsatisfactory, where the respondent was bored by having material read "at" him or her, how the respondent used the various display cards, and which question9 tended to elicit meaningless answers.

With a single exception, the resulting changes were not major and the basic structure of Version IV was retained. The major change was to offer the respondents the explicit opportunity to revise their answers, first after they learned what was required of them by giving their Initial set of values, next, after they were given additional information about what they were actually paying for water (water and air) quality and, finally, when they were presented with a new contingency. The resulting iterative format gave us the series of four WTP amounts described in chapter 1. The advantages of this format are several. (1) Respondents who misunderstood the scenario were offered an unthreatening opportunity to adjust their WTP amounts to reflect their more. comlete understanding of what they were being asked to do. (2) The understandable desire of many respondents to know what they were already paying for waterquality before they gave their WTP amounts could be addressed by

telling them that they would be given this information at a later point in the interview and promising them the opportunity to revise their amounts at that point. Most respondents were sufficiently mollified by this offer to give noninformed WTP amounts. (3) By splitting the sample and informing one half what they were paying for both air and water quality, we could test for the effect of this additional Information. (4) We gained a greater understanding of how respondents' value water quality by being able to analyze the differences between the four data points. (5) The multivalue process enabled respondents to think through the value of the good In a way that a single shot format would not have made possible.

Among the other changes we made at this stags are the following:

- We changed the order of the water quality soenario to place the value card before the payment card. Although seemingly a minor change, it greatly aided the flow of the interview.
- The explanation of the anchors on the payment card was expanded.
- We increased the number of income categories which received separate payment cards by adding a category of $50,000 and above.
- We dropped the map because it did not appear to be needed. This finding surprised everyone who was working on the revision at this point,although there were indications of this finding in the RTI pretest.
- The water quality ladder was included on the reminder card which each respondent was given to use during the sequence of WTP questions. This enhanced the connection in the respondents' minds between the ladder and the amounts they gave under the several conditions.
- We referred to the water quality levels as "goals" C, B and A and 90 identified them on the ladder.
- We added a series of followup questions which were asked of each person who gave a zero answer.
- Several display cards were reformatted, the scenario description was simplifide somewhat, and minor wording changes were made in a number of questions.

Our probes of the zero bidders bears further comment. It has become commonplace in CV studies to ask every zero bidder why they offered that amount in order to determine whether it is a true value or whether it represents a protest against some aspect of the scenario. This information permits the protest zeros to be dropped from tbe analysis as nonresponses. We decided to experiment with a somewhat different format whose goal was to persuade the protest zero bidders to change their mind and express their willingness to pay. Our procedure (questions Y1 to Y11 In the questionnaire), worked as follows. First, we asked the zero bidders if they did so because "that is what it is worth to me (my household)" or because of other reasons. Eighty five percent of the zero bidders said "other reasons. " Second, these 135 people, plus the 90 who gave don't know answers to the WTP questions, were asked up to four followup questions each of which inquired about whether they answered in this way because of a stated reason. If they said no to the first reason, they were asked about the second etc. As soon as a respondent said yes to one of these reasons, a briefcounterargument to the reason was stated after which they were asked if tbey would now be tilling to answer the WTP questions.

This effort must be Judged a failure to some degree. Although almost all the protest zero and "don't know" respondents identified one of the four reasons as characterizing their motivation for giving us a nonresponse to the WTP questions, only 14 of the 217 potential converts were sufficiently convinced by the counterargument to give us valid WTP amounts. For example, when asked: "Did you give this answer (zerooor don't know) because you are (your household is) paying too much in taxesalready and don't want to spend more?", 116 of these respondents said yes. Each of these people were then told :

> "I'd like to remind you that you. . .are already paying some amount for water pollution control in your taxes and prices. It is very

> important to us to learn what value YOU place on achieving the water quality goals when you are given the chance to sake the ohoioe yourself. Would you be willing to answer these questions if I later tell you bow much you are. ..currently paying in taxes and prices snd give you the abanoe to make any changes in your answers you would like to make?" (Question Y4)

Only four agreed to answer tbe WTP questions after this appsal. Either our effort to make persuasive arguments of this kind was a failure or these respondents' rejection of the scenario was so fundamental tbat gentle prodding was doomed to failure. The effort was worth trying as we bad nothing to lose; the respondents were already nonrespondents to the WTP questions. And we did gain 14 additional usable cases.[5]

Together all the changes we made at this 'stage produced an Instrument which was discernibly improved. Row would it work in the field? A revised instrument was pretested by an ORC interviewer who conducted twenty field interviews in the Princeton and Trenton inner city areas. On the basis of her favorable report Version V (see appendix A for the text) was prepared and administered as the final version. The ORC staff closely monitored the field interviewing so that any further problems would he quickly identified and any necessary remedial action could Instituted. No generic problems were identified; interviewers reported the instrument worked smoothly.

---

5. Our procedure also gave us valuable Information in the form of written comments describing the respondents' reasons why they gave a zero bid. It is possible that more effective argments for a future version of this questionnairecould be developed on the basis of these coments.

5. THE PAYMENT CARD ELICITATION METHOD

Of all the design decisions which must be made in designing a contingent valuation scenario, the choice of the elicitation procedure is both one of the most crucial and the most controversial. It is crucial because the elicitation procedure defines the respondent's task at the point when the respondent values the good. The least-structured form is the question, "How much are you willing to pay for X?" Because respondents tend to have difficulty answering this type of open ended question when confronted with the unfamiliar and somewhat intimidating task of valuing a public good, [1] alternative procedures have been developed which assist the respondent by providing more structure. The bidding game asks for a yes or no answer to a given amount which is then iterated up or down until the respondent gives the opposite answer; the variable offer approach uses a large number of starting points only one of which is offered to each respondent on a take it or leave it basis with no iteration; the contingent ranking approach has respondents order combinations of amenity level and WTP amounts; the check list procedure asks the respondent to choose one of a set of payment ranges; the payment card procedure provides a menu of specific

---

1. For example, In Seller, Stoll and Chavas'(1983) comparison of the open ended question and the variable offer approaches, 25 percent of those who received the open ended question said they could not provide an accurate answer compared with 9 percent of those in the variable offer treatment.

2. Bishop and Heberlein (1979, 1980) developed this approach which has several desirable features and is particularly suitable to mail questionnaires.

amounts for the respondent to choose from;and the anchored payment card format offers additional information about what people like the respondent are paying for other public goods.

The elicitation procedure choice is controversial because researchers hold different views about which technique best meets the criteria for a valid and reliable elicitation method. The ideal procedure should: 1) result in the highest possible number of usable responses, 2) obtain the maximum willingness to pay, 3) minimize the possibility of bias from Implied value cues, 4) minimize the variance of the elicited WTP amounts, 5) not require more interviews than the open ended question and, 6) be easily administered by the interviewer (or self-administered, if a mail questionnaire procedure is used). This is a formidable set of criteria which Inevitably requires tradeoffs.

When we began work on this study in 1979,the bidding game was the prevailing elicitation method for contingent valuation studies. We came to the conclusion that this technique was not appropriate for a study such as ours where the respondents were likely to be initially uncertain about their values and therefore prone to rely on cues provided by the elicitation method. Our substitute, the anchored payment card, was designed to assist the respondents' valuation effort without itself implying a value. Since this method was novel, we conducted tests to determine its properties and to see If it's anchor feature biased the answers in any way. Other studies subsequently have compared the payment card with one or more of the other elicitation techniques. In this chapter we review the considerations which led us to reject the bidding game approach, describe the anchored payment card technique, and assess the degree to which It meets the above criteria.[3]

---

3. For a comparative evaluation of all the elicitation methods, including those we do not discuss here, see Mitchell and Carson (1984).